

161444

Shears, Beverly

From: Devi, Sarvamangala  
Sent: Wednesday, July 27, 2005 3:59 PM  
To: Shears, Beverly  
Subject: 10/070,882

Beverly:

In application 10/070,882, please perform a sequence search for SEQ ID NO: 2 in commercial and pending databases.

Thanx.

S. DEVI, Ph.D.  
Primary Examiner  
AU 1645  
Rems - 3C18

UA  
Need pending

1

Date completed:

Searcher:

Beverly e 2528

Terminal time:

Elapsed time:

CPU time:

Total time:

Number of Searches:

Number of Databases:

Search Site

STIC

CM-1

Pre-S

Type of Search

N.A. Sequence

A.A. Sequence

Structure

Bibliographic

Vendors

IG

STN

Dialog

APS

Geninfo

SDC

DARC/Questel

Other

CGN

**This Page Blank (uspto)**



# **STIC Search Report**

## **Biotech-Chem Library**

STIC Database Tracking Number 161444

**TO: Sarvamangala Devi**  
**Art Unit: 1645**  
**Location: REM 3C18**  
**Serial Number: 10/070882**

**Tuesday, August 09, 2005**

**From: Beverly Shears**  
**Location: Biotech-Chem Library**  
**REM 1A54**  
**Phone: 571-272-2528**  
**beverly.shears@uspto.gov**

### **Search Notes**

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

Off nucleic - nucleic search, using sw model

Run on: August 3, 2005, 21:30:55 ; Search time 7070 Seconds

(without alignments)  
748.364 Million cell updates/sec

Title: US-10-070-882A-2

Perfect score: 139

Sequence: 1 gtgactctgtgcgcgaact.....ctgtcttcatgttaacaca 139

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: gb\_esc1:\*

2: gb\_esc2:\*

3: gb\_hic:\*

4: gb\_esc3:\*

5: gb\_esc4:\*

6: gb\_esc5:\*

7: gb\_esc6:\*

8: gb\_gsa1:\*

9: gb\_gsa2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	34.8	25.0	3932	3	AK082454 Mus muscu
2	34.4	24.7	414	6	BY642267 BY642267
3	33.6	24.2	204	1	AV231406 AV231406
4	33.6	24.2	274	2	BB775700 BB775700
5	33.6	24.2	318	2	BB094276 BB094276
6	33.6	24.2	335	2	BB222962 BB222962
7	33.6	24.2	448	5	BY439440 BY439440
8	33.6	24.2	570	8	AZ435636 IM0222H14
9	33.6	24.2	590	2	BE131381 L48-1355T
10	33.6	24.2	605	4	BM658097 MCR05H01
11	33.6	24.2	625	6	CA835622 MCR039B06
12	33.6	24.2	633	6	CA834105 MCR028H02
13	33.6	24.2	674	4	BM301497 MCR046F12
14	33.6	24.2	699	9	BM301422 MCR045F12
15	33.6	24.2	746	9	CR303205 Medicago
16	33.6	24.2	863	3	BE035174 MCR22B09
17	33.6	24.2	866	9	CR308064 Medicago
18	33.6	24.2	889	9	CG952370 MCR039B06
19	33.6	24.2	1192	2	BE037111 MCR039B06
20	33.4	24.0	521	8	AG0830244 HS 4825_B
21	33.4	24.0	728	9	AG308530 Mus muscu
22	33.4	24.0	957	6	CA791309 AGRNCCURT
23	33.2	23.9	205	2	BB004186 BB004186
24	33	23.7	467	8	BZ915818 CR240_59A

25	32.8	23.6	206	1	AV336201 AV336201
26	32.8	23.6	690	8	BZ035953 BZ035953
27	32.4	23.3	252	2	BB230680 BB230680
28	32.4	23.3	270	2	BB118576 BB118576
29	32.4	23.3	533	5	BY480819 BY480819
30	32.4	23.3	906	8	AQ781019 HS_3169_B
31	32.2	23.2	275	1	AI1414763 ma59d02.x
32	32.2	23.2	459	2	BE289932 601089159
33	32.2	23.2	579	1	AI981576 Pat. PK006
34	32.2	23.2	741	9	CC537179 CC537179
35	32.2	23.2	769	8	AQ869785 AQ869785
36	32	23.0	239	1	AA686653 AA686653
37	32	23.0	274	2	BB364524 BB364524
38	32	23.0	320	2	BB464866 BB464866
39	32	23.0	391	5	BY468249 BY468249
40	32	23.0	490	8	AQ430313 HS_5093_A
41	32	23.0	501	8	BH002081 BH002081
42	32	23.0	611	7	CO040889 CO040889
43	32	23.0	706	8	BH830539 BH830539
44	32	23.0	745	9	AG479014 AG479014
45	31.8	22.9	175	9	CE320783 CE320783

## ALIGNMENTS

RESULT 1  
AK082454/c  
LOCUS  
DEFINITION  
Mus musculus 0 day neonate cerebellum cDNA, RIKEN full-length  
enriched library, clone: C230052F14 product: unclassified, full  
insert sequence.

ACCESSION  
AK082454  
VERSION  
AK082454.1 GI:26100683  
KEYWORDS  
HTC, CAP trapper.  
SOURCE  
Mus musculus (house mouse)  
ORGANISM  
Mus musculus

REFERENCE  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
High-efficiency full-length cDNA cloning  
Meth. Enzymol. 303, 19-44 (1999)  
JOURNAL  
99279253  
MEDLINE  
10349636  
PUBMED

REFERENCE  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new genes  
Genome Res. 10 (10), 1617-1630 (2000)  
JOURNAL  
20499374  
MEDLINE  
11042159  
PUBMED

## REFERENCES

Shibata, K., Itoh, M., Aizawa, K., Nagaoaka, S., Sasaki, N., Carninci, P.,  
Kono, H., Akiyama, J., Nishi, K., Katsunuma, T., Tachino, H., Itoh, M.,  
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, H., Harada, A.,  
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,  
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M.,  
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Maruyama, S., Kawai, J.,  
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system-384-format  
sequencing pipeline with 384 multichannel sequencer  
Genome Res. 10 (11), 1757-1771 (2000)  
JOURNAL  
20530913  
MEDLINE  
11076861  
PUBMED

## REFERENCES

The RIKEN Genome Exploration Research Group Phase II Team and the  
FANTOM Consortium.  
Functional annotation of a full-length mouse cDNA collection  
Nature 409, 685-690 (2001)  
JOURNAL  
5  
PUBMED  
The FANTOM Consortium and the RIKEN Genome Exploration Research



## FEATURES

Location/Qualifiers

```
1. 414
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="K430350A22"
/tissue_type="visual cortex"
/clone_1ib="RIKEN full-length enriched, visual cortex"
```

## ORIGIN

Query Match 24.7%; Score 34.4; DB 6; Length 414;  
Best Local Similarity 57.4%; Pred. No. 12; Mismatches 46; Indels 0; Gaps 0;  
Matches 62; Conservative 0;

Qy 31 CTGCTCACCCCTCTTTCTTTCAGAAAGAGGTGACTATTGTCGTTTAACTGTTT 90  
214 CTGCTCACGCTCTTCTTCTTGTGAGCTGTGCTCCACGTTCAATCTTCACTGTTT 273  
Db 91 ATCCCAAGACCATATCAACGCTAGACTGTTCTTATTGTTAAC 138  
274 TTCTTCAGACACACCCACATGCTTGACAGTCTTGTTCAGCGC 321

RESULT 3 AV231406 204 bp mRNA linear EST 15-NOV-2001  
AV231406 RIKEN full-length enriched, 0 day neonate skin Mus  
LOCUS musculus cDNA clone 4632406L08 3', mRNA sequence.  
DEFINITION AV231406.1 GI:6183921

ACCESSION AV231406.1 GI:6183921  
VERSION EST.  
KEYWORDS Mus musculus (house mouse)  
SOURCE Mus musculus  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 204)  
AUTHORS Kono, H., Aizawa, K., Akahira, S., Akiyama, J., Carninci, P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Horii, F., Ishii, Y., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, T., Kai, C., Kawai, J., Kikuchi, N., Kojima, Y., Koya, S., Kusakabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y., Owa, C., Ozawa, Y., Saito, H., Sano, M., Sato, K., Shibata, K., Shibata, Y., Shigemoto, Y., Shiraki, T., Sogabe, Y., Suganara, Y., Suzuki, H., Suzuki, H., Takahashi, F., Tateo, M., Tomimaga, N., Tsunoda, Y., Watabiki, A., Watanabe, S., Yamamura, T., Yasunishi, A., Yokota, T., Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.  
RIKEN Mouse ESTs (Kono, H., et al. 1999)  
Unpublished (1999)  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-resgsc.riken.jp, URL: http://genome-gsc.riken.jp/  
Sasaki, M., Izawa, M., Watabiki, M., Ozawa, K., Tanaka, T., Yoneda, Y., Matsura, S., Carninci, P., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.  
Transcriptional sequencing: A method for DNA sequencing using RNA polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)  
Ito, M., Katsunari, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J., Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.  
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)  
Carninci, P. and Hayashizaki, Y.  
High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)  
Please visit our web site (http://genome-gsc.riken.go.jp) for further details.  
Location/Qualifiers

## FEATURES

## source

```
1. 204
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="4632406L08"
/sex="mixed"
/tissue_type="skin"
/dev_stage="0 day neonate"
/lab_host="DH10B"
/clone_1ib="RIKEN full-length enriched, 0 day neonate skin"
```

## ORIGIN

Query Match 24.2%; Score 33.6; DB 1; Length 204;  
Best Local Similarity 57.7%; Pred. No. 18; Mismatches 44; Indels 0; Gaps 0;  
Matches 60; Conservative 0;

Qy 31 CTGCTCACCCCTCTTTCTTTCAGAAAGAGGTGACTATTGTCGTTTAACTGTTT 90  
10 CTGCTCACGCTCTTCTTCTTGTGAGCTGTGCTCCACGTTCAATCTTCACTGTTT 69  
Db 91 ATCCCAAGACCATATCAACGCTAGACTGTTCTTATTGTTTAA 134  
70 TTCTTCAGACACACCCCAATGCTTGACAGTCTTGTTCAGCGC 113

RESULT 4 BB775700 274 bp mRNA linear EST 08-JUL-2003  
BB775700 RIKEN full-length enriched, RCB-0559 K-1. F1 cDNA Mus  
LOCUS musculus cDNA clone G430026L09 3', mRNA sequence.  
DEFINITION BB775700.1 GI:16936380

ACCESSION BB775700.1 GI:16936380  
VERSION EST.  
KEYWORDS Mus musculus (house mouse)  
SOURCE Mus musculus  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 274)  
AUTHORS Akimura, T., Aikawa, T., Carninci, P., Furuno, M., Hanagaki, T., Hayatsu, N., Hiramoto, K., Hirozane, T., Hirozane, T., Imotani, K., Ishii, Y., Ito, M., Kawai, J., Kojima, Y., Kono, H., Kouda, M., Matsuyama, T., Nakamura, M., Nishi, K., Nomura, K., Nunasaki, R., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagawa, A., Takahashi, F., Takaku-Kakihira, S., Tanaka, T., Tomaru, A., Toya, T., Watabiki, A., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.  
RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura, T., et al. 2001)  
Unpublished (2001)  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)

1-7-22	Sheniro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel:	81-45-503-9222
Fax:	81-45-503-9216
Email:	genome-res@gsic.riken.jp, URL: http://genome.gsic.riken.jp/
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Komoto, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.	
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)	
genes. Genome Res. 10 (10), 1617-1630 (2000)	
wagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matsubara, S., Kawai, J., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.	
RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)	
Komoto, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y., and Hayashizaki, Y.	
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)	
Please visit our web site ( <a href="http://genome.gsic.riken.go.jp">http://genome.gsic.riken.go.jp</a> ) for further details.	
e mouse tissues.	
location/Qualifiers	
1..274	
/organism="Mus musculus"	
/mol_type="mRNA"	
/strain="DDD"	
/db_xref="taxon:10090"	
/clone="G430026109"	
/cell_line="RCB-0559 K-1.fl"	
/clone_idb="RIKEN full-length enriched, RCB-0559 K-1 . fl cDNA"	
/note="Organ: spleen"	
ORIGIN	
Query Match	24.2%; Score 33.6; DB 2; Length 274;
Best Local Similarity	57.7%; Pred. No. 19;
Matches	60; Conservative 0; Mismatches 44; Indels 0; Gaps 0;
Cy	31 CTGGCTCAACCCCTTTTCTTCGAAAGAGGGTGACTATTGTCGTATTATTAATGTTT 90
Db	80 CTGCTCCAGCTCTTCTCTCTTCGACCTGCTGTCCTCCACTTTCAGATCTTTCACGTITT 139
Cy	91 ATCCCAAGACCATTAATCAACGCTAGACTGTTCTTATTGTTA 134
Db	140 TTCTTCAGAGACACCCACACATTTGTTGACAGTCTTTTTC 183
RESULT 5	
BB094276	318 bp mRNA linear EST 26-JUN-2000
LOCUS	BB094276
DEFINITION	BB094276 RIKEN full-length enriched, 12 days embryo, embryonic body between diaphragm region and neck Mus musculus cDNA clone
ACCESSION	BB094276
VERSION	BB094276.1
KEYWORDS	EST.
SOURCE	Mus musculus (house mouse)
ORGANISM	Mus musculus
REFERENCE	Eukaryotic; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS	1 (bases 1 to 318)
	Komoto, H., Aizawa, K., Akahira, S., Akiyama, J., Arakawa, T., Carninci, P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Iizawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N., Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurahara, C., Kusakabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y., Oho, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata, Y., Shigenoto, Y., Shingawa, A., Shitaki, T., Sugabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Tagawa, A.,

TITLE  
JOURNAL  
COMMENT

RIKEN Mouse ES/ES (Kono, H., et al.)  
Unpublished (2000)  
Contact: Yoshinori Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@sc.riken.jp, URL: http://genome.gsc.riken.jp/  
Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoaka, S.,  
Sasaki, N., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.,  
Thermolabile and thermolabile activation of thermolabile enzymes by  
trehalose and its application for the synthesis of full length  
cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)  
Itoh, M., Kikunishi, T., Akiyama, J., Shibata, K., Iwase, M., Kawai, J.,  
Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M.,  
Okazaki, Y., and Hayashizaki, Y.,  
Automated filtration-based high-throughput plasmid preparation  
system. Genome Res. 9 (5), 463-470 (1999)  
Carninci, P. and Hayashizaki, Y.,  
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,  
19-44 (1999)  
Please visit our web site (<http://genome.riken.go.jp>) for  
further details.

Location/Qualifiers

1. 318

/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="9430045021"  
/tissue\_type="embryonic body between diaphragm region and  
neck"  
/dev\_stage="12 days embryo"  
/lab\_host="DH10B"  
/clone\_1db="RIKEN full-length enriched, 12 days embryo,  
embryonic body between diaphragm region and neck"  
/note="Site 1: Sali; Site 2: BamHI; cDNA library was  
prepared and sequenced in Mouse Genome Encyclopedia  
Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in  
RIKEN. Division of Experimental Animal Research in Riken  
contributed to prepare mouse tissues. 1st strand cDNA was  
primed with a primer [5'  
GAGGAGAGAGAGATCCAGACGCTCTTTTCTTTTCTTTTNN 3'], cDNA was  
prepared by using trehalose thermo-activated reverse  
transcriptase and subsequently enriched for full-length by  
cap-trapper. cDNA went through one round of normalization  
to Rot = 20.0 and subtraction to Rot = 370.0. Second  
strand cDNA was prepared with the primer adapter of  
sequence [5' GAGGAGAGATTCGAGGTATTAATTAATATCCCCCCCCCC  
3']. cDNA was cleaved with XhoI and BamHI."

ORIGIN

Query Match 24.2%; Score 33.6; DB 2; Length 318;  
Best Local Similarity 57.7%; Pred. No. 20;  
Matches 60; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

31 CTGCTCAACCTCTTTTCTTCAAGAAAGGTGACATTTGTCTGCTTATTAACCTGTT 90  
|||||  
122 CTGCTCCAGAGATCTTCTTTTCTGAGCTGTGCTCCCACTGTTGAGATTTTTCACCTGTT 181  
|||||

91 ATCCCAAGACCATATCAACCTAGACCTGTCTTATTTGTTA 134  
|||||

DB 182 TTCTTCAGGACACCCACACATTCCTGACAGCTCTGTGTGCA 225  
|||||



Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Kono, H., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Waki, K., Watanabe, A., Muramatsu, M. and Hayashizaki, Y. Direct SubMISSION Computational Analysis of Full-length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)

Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)

RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

# FEATURES

## source

```
1..448
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="1920162D12"
/clone_lib="RIKEN full-length enriched, pooled tissues, 16 days embryo, etc."
/notes="pooled tissues: (dev_stage=16 days embryo, tissue_type=heart, sex=mix), (dev_stage=16 days embryo, tissue_type=kidney, sex=mix), (dev_stage=17 days embryo, tissue_type=heart, sex=mix), (dev_stage=17 days embryo, tissue_type=stomach, sex=mix), (dev_stage=17 days embryo, tissue_type=kidney, sex=mix), (dev_stage=17 days pregnant, adult, tissue_type=emlion, sex=female), (dev_stage=13 days embryo, tissue_type=liver, sex=mix)"
```

# ORIGIN

Query Match 24.2%; Score 33.6; DB 5; Length 448;  
 Best Local Similarity 61.4%; Pred. No. 21;  
 Matches 54; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

```
QY 47 TCTTCAGAAAGAGGCTGACTATTGTCGTTATTACTGTTATCCCAAGACCAT 106
    |||||
DB 262 TCTTCGAGCTGTGGGTCCGCTGCAGATCTCTCAGTGTTCCTTCAGAGCACC 321
    |||||
QY 107 AATCAAGCTAGACTGTTCTTATTGTTA 134
    |||||
DB 322 ACACATGCTTGACAGCTCTTGTCGA 349
```

RESULT 8  
 AZ435636 570 bp DNA linear GSS 03-OCT-2000  
 LOCUS IM0222H14R Mouse 10kb plasmid UGCGIM library Mus musculus genomic  
 DEFINITION Clone UGCGIM0222H14 R, genomic survey sequence.  
 ACCESSION AZ435636  
 VERSION AZ435636.1 GI:10559649  
 KEYWORDS GSS.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
 AUTHORS Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamli, C., Islem, H., Longacre, S., Mamoud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A. and Wright, D., Weiss, R.  
 TITLE Mouse whole genome scaffolding with paired end reads from 10kb

# JOURNAL COMMENT

plasmid inserts  
 Unpublished (2000)  
 Contact: Robert B. Weiss  
 University of Utah Genome Center  
 University of Utah  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT  
 84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: ddunne@genetics.utah.edu  
 Insert Length: 10000 Std Error: 0.00  
 Plate: 0222 row: H column: 14  
 Seq primer: CACACGAGAAACGCTATGACC  
 Class: plasmid ends  
 High quality sequence stop: 570.

# FEATURES

## source

```
1..570
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UGCGIM0222H14"
/sex="Male"
/lab_host="B. CoII strain X110-Gold, T1-resistant, F-"  

/clone_lib="Mouse 10kb plasmid UGCGIM library"  

/notes="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PWD42 (914732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli X110-Gold (Stratagene) cells and selected for ampicillin resistance."
```

# ORIGIN

Query Match 24.2%; Score 33.6; DB 8; Length 570;  
 Best Local Similarity 63.8%; Pred. No. 22;  
 Matches 51; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

```
QY 1 GTGACTCTGTGAGCAACTTAATATGCGCTGACCCCTCTTCTTTCAGAAAGAG 60
    |||||
DB 216 GTGACTTTAGTAAAGCTAATTATTTCTCCATTCTCATGCTTACTATTAAACAG 157
    |||||
QY 61 GTGACTATTGTCGTGTTA 80
    |||||
DB 156 GAGAAATGTTTGTTCCTTA 137
```

RESULT 9  
 BE131381 590 bp mRNA linear EST 20-FEB-2001  
 LOCUS I48-1355T3 Ice plant lambda Uni-Zap XR expression library, 48 hours  
 DEFINITION NaCl treatment Mesembryanthemum crystallinum cDNA clone I48-1355,  
 mRNA sequence.  
 ACCESSION BE131381  
 VERSION BE131381.1 GI:8578744  
 KEYWORDS EST.  
 SOURCE Mesembryanthemum crystallinum (common iceplant)  
 ORGANISM Mesembryanthemum crystallinum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Caryophyllales; Alismaceae; Mesembryanthemum.

## REFERENCE 1 (bases 1 to 590)

AUTHORS Cushman, J.C.  
TITLE An expressed sequence tag database for the common ice plant,  
Mesembryanthemum crystallinum

JOURNAL Unpublished (1997)  
COMMENT Contact: Cushman JC  
Department of Biochemistry  
University of Nevada  
MS200, Reno, NV 89557-0014, USA  
Tel: 775-784-1918  
Fax: 775-784-1650  
Email: jcushman@unr.edu

PCR Primers  
FORWARD: T7  
BACKWARD: T3  
Plate: L48-14 row: E column: 7  
Seq primer: T3  
High quality sequence stop: 350  
POLYA-No.  
Location/Qualifiers

FEATURES  
source 1..590  
/organism="Mesembryanthemum crystallinum"  
/mol\_type="mRNA"  
/db\_xref="taxon:3544"  
/clone="L48-1355"  
/cissue\_type="leaf, 48 h 0.4M NaCl"  
/dev\_stage="Six week old"  
/clone\_1ib="Ice plant Lambda Uni-Zap XR expression  
library, 48 hours NaCl treatment"  
/note="Vector: Lambda Uni-Zap XR, Bluescript SK-; Site\_1:  
EcoRI; Site\_2: XhoI"

## ORIGIN

Query Match 24.2%; Score 33.6; DB 2; Length 590;  
Best Local Similarity 61.4%; Pred. No. 22;  
Matches 54; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 44 TTTTCTCGAAGAGGGGACTATTGTCGTGTTATTAAGCTTTATCCCAAGCAC 103  
|||||  
DB 147 TTTTCATCAAGACAGGTGACTTGGGTTGAATTTCTATCCCTAATCAACAAATCAC 206  
|||||

QY 104 CATATCAACGCTAGACTGTTCTTATG 131  
|||||  
DB 207 CTGAAAAAACAGTGGAGTTGTTTGG 234  
|||||

RESULT 10 605 bp mRNA linear EST 26-FEB-2002  
BM658097  
LOCUS MCR059H01.69977 Ice plant Lambda Uni-Zap XR expression library, 48  
DEFINITION hours NaCl treatment prescreened for removal of highly abundant  
transcripts Mesembryanthemum crystallinum cDNA clone MCR059H01.5,  
mRNA sequence.

ACCESSION BM658097 GI:18957608  
VERSION BM658097.1  
KEYWORDS EST.  
SOURCE Mesembryanthemum crystallinum (common iceplant)  
ORGANISM Mesembryanthemum crystallinum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Caryophyllales; Aizoaceae; Mesembryanthemum.

REFERENCE 1 (bases 1 to 605)  
Cushman, J.C.  
An expressed sequence tag database for the common ice plant,  
Mesembryanthemum crystallinum

JOURNAL Unpublished (1997)  
COMMENT Contact: Cushman JC  
Department of Biochemistry  
University of Nevada  
MS200, Reno, NV 89557-0014, USA  
Tel: 775-784-1918  
Fax: 775-784-1650  
Email: jcushman@unr.edu

PCR Primers  
FORWARD: T3 20mer  
BACKWARD: T7 21mer  
Plate: 059 row: H column: 01  
Seq primer: T3 20mer  
High quality sequence stop: 605.  
Location/Qualifiers

FEATURES  
source 1..605  
/organism="Mesembryanthemum crystallinum"  
/mol\_type="mRNA"  
/db\_xref="taxon:3544"  
/clone="MCR059H01"  
/cissue\_type="leaf"  
/dev\_stage="six-week-old"  
/clone\_1ib="Ice plant Lambda Uni-Zap XR expression  
library, 48 hours NaCl treatment prescreened for removal  
of highly abundant transcripts"  
/note="Vector: Lambda Uni-Zap XR, Bluescript SK-; Site\_1:  
EcoRI; Site\_2: XhoI"

## ORIGIN

Query Match 24.2%; Score 33.6; DB 4; Length 605;  
Best Local Similarity 61.4%; Pred. No. 22;  
Matches 54; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 44 TTTTCTCGAAGAGGGGACTATTGTCGTGTTATTAAGCTTTATCCCAAGCAC 103  
|||||  
DB 153 TTTTCATCAAGACAGGTGACTTGGGTTGAATTTCTATCCCTAATCAACAAATCAC 212  
|||||

QY 104 CATATCAACGCTAGACTGTTCTTATG 131  
|||||  
DB 213 CTGAAAAAACAGTGGAGTTGTTTGG 240  
|||||

RESULT 11 625 bp mRNA linear EST 12-DEC-2002  
CA835622  
LOCUS MCS039B06.160752 Ice plant Lambda Uni-Zap XR expression library, 5  
DEFINITION days 0.5 M NaCl treatment, Crasulacean acid metabolism, phase I (2  
AM). Mesembryanthemum crystallinum cDNA clone MCS039B06.5, mRNA  
sequence.

ACCESSION CA835622  
VERSION CA835622.1 GI:26563387  
KEYWORDS EST.  
SOURCE Mesembryanthemum crystallinum (common iceplant)  
ORGANISM Mesembryanthemum crystallinum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Caryophyllales; Aizoaceae; Mesembryanthemum.

REFERENCE 1 (bases 1 to 625)  
Cushman, J.C.  
An expressed sequence tag database for the common ice plant,  
Mesembryanthemum crystallinum

JOURNAL Unpublished (1997)  
COMMENT Contact: Cushman JC  
Department of Biochemistry  
University of Nevada  
MS200, Reno, NV 89557-0014, USA  
Tel: 775-784-1918  
Fax: 775-784-1650  
Email: jcushman@unr.edu

PCR Primers  
FORWARD: T3 20mer  
BACKWARD: T7 21mer  
Plate: 039 row: B column: 06  
Seq primer: T3 20mer  
High quality sequence stop: 625.  
Location/Qualifiers

FEATURES  
source 1..625  
/organism="Mesembryanthemum crystallinum"  
/mol\_type="mRNA"  
/db\_xref="taxon:3544"  
/clone="MCS039B06"

/tissue\_type="leaf"  
/dev\_stage="five-week-old"  
/clone\_lib="Ice plant lambda Uni-Zap XR expression  
library", 5 days 0.5 M NaCl treatment, Crassulacean acid  
metabolism, phase I (2 AM)."  
/note="Vector: Lambda Uni-Zap XR, Bluescript SK-, Site\_1:  
EcoRI; Site\_2: XhoI; Library construction was performed  
according to Stratagene's recommended protocol for the  
Lambda UniZapXR vector and cDNA synthesis kit."

## ORIGIN

Query Match 24.2%; Score 33.6; DB 6; Length 625;  
Best Local Similarity 61.4%; Pred. No. 22;  
Matches 54; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 44 TTTTCTGAGAAAGAGGCTATTGTCGTTTAACTGTTATCCCAAGCAC 103  
|||||  
DB 156 TTTTCATCACAAGAGGAGCTTGGGTTGAAATCTTATCCCTAATCAACAATCAC 215  
|||||  
QY 104 CATATCAACGCTAGACTGTTCTTATG 131  
|||||  
DB 216 CTGAAAAAACAAGTGGAGTTGTTTGG 243  
|||||

## RESULT 12

CA834105 633 bp mRNA linear EST 12-DEC-2002  
LOCUS CA834105  
DEFINITION MCS028H02\_153488 Ice plant lambda Uni-Zap XR expression library, 5  
days 0.5 M NaCl treatment, Crassulacean acid metabolism, phase I (2  
AM). Mesembryanthemum crystallinum cDNA clone MCS028H02\_5, mRNA  
sequence.

ACCESSION CA834105  
VERSION CA834105.1 GI:26561870  
KEYWORDS EST

SOURCE Mesembryanthemum crystallinum (common iceplant)  
ORGANISM Mesembryanthemum crystallinum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Caryophyllales; Alzooceae; Mesembryanthemum.

REFERENCE 1 (bases 1 to 633)  
Cushman, J.C.

AUTORS An expressed sequence tag database for the common ice plant,  
TITLE Mesembryanthemum crystallinum  
JOURNAL Unpublished (1997)

## COMMENT

Contact: Cushman JC  
Department of Biochemistry  
University of Nevada  
MS200, Reno, NV 89557-0014, USA  
Tel: 775-784-1918  
Fax: 775-784-1650  
Email: jcushman@unr.edu  
PCR Primers  
FORWARD: T3 20mer  
BACKWARD: T7 21mer  
Plate: 028 row: H column: 02  
Seq primer: T3 20mer  
High quality sequence stop: 633.

## FEATURES

source

1..633  
/organism="Mesembryanthemum crystallinum"  
/mol\_type="mRNA"  
/db\_xref="taxon:3544"  
/clone="MCS028H02"  
/tissue\_type="leaf"  
/dev\_stage="five-week-old"  
/clone\_lib="Ice plant lambda Uni-Zap XR expression  
library, 5 days 0.5 M NaCl treatment, Crassulacean acid  
metabolism, phase I (2 AM)."  
/note="Vector: Lambda Uni-Zap XR, Bluescript SK-, Site\_1:  
EcoRI; Site\_2: XhoI; Library construction was performed  
according to Stratagene's recommended protocol for the  
Lambda UniZapXR vector and cDNA synthesis kit."

## ORIGIN

Query Match 24.2%; Score 33.6; DB 6; Length 633;  
Best Local Similarity 61.4%; Pred. No. 22;  
Matches 54; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 44 TTTTCTGAGAAAGAGGCTATTGTCGTTTAACTGTTATCCCAAGCAC 103  
|||||  
DB 165 TTTTCATCACAAGAGGAGCTTGGGTTGAAATCTTATCCCTAATCAACAATCAC 224  
|||||  
QY 104 CATATCAACGCTAGACTGTTCTTATG 131  
|||||  
DB 225 CTGAAAAAACAAGTGGAGTTGTTTGG 252  
|||||

## RESULT 13

BM301497 674 bp mRNA linear EST 22-JAN-2002  
LOCUS BM301497  
DEFINITION MCR046F12\_26036 Ice plant lambda Uni-Zap XR expression library, 48  
hours NaCl treatment prescreened for removal of highly abundant  
transcripts Mesembryanthemum crystallinum cDNA clone MCR046F12\_5,  
mRNA sequence.

ACCESSION BM301497  
VERSION BM301497.1 GI:18023872  
KEYWORDS EST

SOURCE Mesembryanthemum crystallinum (common iceplant)  
ORGANISM Mesembryanthemum crystallinum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Caryophyllales; Alzooceae; Mesembryanthemum.

REFERENCE 1 (bases 1 to 674)  
Cushman, J.C.

AUTORS An expressed sequence tag database for the common ice plant,  
TITLE Mesembryanthemum crystallinum  
JOURNAL Unpublished (1997)

## COMMENT

Contact: Cushman JC  
Department of Biochemistry  
University of Nevada  
MS200, Reno, NV 89557-0014, USA  
Tel: 775-784-1918  
Fax: 775-784-1650  
Email: jcushman@unr.edu  
PCR Primers  
FORWARD: T3 20mer  
BACKWARD: T7 21mer  
Plate: 046 row: F column: 12  
Seq primer: T3 20mer  
High quality sequence stop: 674.

## FEATURES

source

1..674  
/organism="Mesembryanthemum crystallinum"  
/mol\_type="mRNA"  
/db\_xref="taxon:3544"  
/clone="MCR046F12"  
/tissue\_type="leaf"  
/dev\_stage="six-week-old"  
/clone\_lib="Ice plant lambda Uni-Zap XR expression  
library, 48 hours NaCl treatment prescreened for removal  
of highly abundant transcripts"  
/note="Vector: Lambda Uni-Zap XR, Bluescript SK-, Site\_1:  
EcoRI; Site\_2: XhoI"

## ORIGIN

Query Match 24.2%; Score 33.6; DB 4; Length 674;  
Best Local Similarity 61.4%; Pred. No. 22;  
Matches 54; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 44 TTTTCTGAGAAAGAGGCTATTGTCGTTTAACTGTTATCCCAAGCAC 103  
|||||  
DB 156 TTTTCATCACAAGAGGAGCTTGGGTTGAAATCTTATCCCTAATCAACAATCAC 215  
|||||  
QY 104 CATATCAACGCTAGACTGTTCTTATG 131  
|||||  
DB 216 CTGAAAAAACAAGTGGAGTTGTTTGG 243  
|||||

RESULT 14  
BM301422 699 bp mRNA linear EST 22-JAN-2002  
LOCUS MCR045f12.25886 Ice plant Lambda Uni-Zap XR expression library, 48  
DEFINITION hours NaCl treatment prescreened for removal of highly abundant  
transcripts Mesembryanthemum crystallinum cDNA clone MCR045f12.5,  
mRNA sequence.  
ACCESSION BM301422  
VERSION BM301422.1 GI:18023797  
KEYWORDS  
SOURCE Mesembryanthemum crystallinum (common iceplant)  
ORGANISM Mesembryanthemum crystallinum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Caryophyllales; Aizoaceae; Mesembryanthemum.  
1 (bases 1 to 699)  
Cushman, J.C.  
REFERENCE An expressed sequence tag database for the common ice plant,  
AUTHORS Mesembryanthemum crystallinum  
TITLE Mesembryanthemum crystallinum  
JOURNAL Unpublished (1997)  
COMMENT Contact: Cushman JC  
Department of Biochemistry  
University of Nevada  
MS200, Reno, NV 89557-0014, USA  
Tel: 775-784-1918  
Fax: 775-784-1650  
Email: jcushman@unr.edu  
PCR Primers  
FORWARD: T3 20mer  
BACKWARD: T7 21mer  
Plate: 045 row: F column: 12  
Seq primer: T3 20mer  
High quality sequence stop: 699.  
Location/Qualifiers  
1..699  
/organism="Mesembryanthemum crystallinum"  
/mol\_type="mRNA"  
/db\_xref="taxon:3544"  
/clone="MCR045f12"  
/issue\_type="leaf"  
/dev\_stage="six-week-old"  
/clone\_1lb="Ice plant Lambda Uni-Zap XR expression  
library, 48 hours NaCl treatment prescreened for removal  
of highly abundant transcripts"  
/note="Vector: lambda Uni-Zap XR, Bluescript SK-, Site\_1:  
EcoRI; Site\_2: XhoI"

ORIGIN  
Query Match 24.2%; Score 33.6; DB 4; Length 699;  
Best Local Similarity 61.4%; Pred. No. 23;  
Matches 54; Conservative 0; Mismatches 34; Indels 0; Gaps 0;  
QY 44 TTTTCTTCGAAAGAGCGTACTATTGCTGTTATTATCTGTTATCCCAAGAC 103  
|||||  
DB 156 TTTTCATCAGAAAGCGAGCTTGCTGGAATCTTATCCCTATCAACAATCAC 215  
QY 104 CATATCAAGCGCTAGACTGTTCTTATTG 131  
|||||  
DB 216 CTGAAAAAACAAGTGGAGTTGTTTG 243

RESULT 15  
CR303205 746 bp DNA linear GSS 28-FEB-2004  
LOCUS CR303205  
DEFINITION Medicago truncatula BAC ends cultivar Jemalong A17 of Medicago  
truncatula, genomic survey sequence.  
ACCESSION CR303205  
VERSION CR303205.1 GI:44708225  
KEYWORDS  
SOURCE Medicago truncatula (barrel medic)  
ORGANISM Medicago truncatula

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;  
Medicago.  
1 (bases 1 to 746)  
AUTHORS Genoscope.  
REFERENCE Direct Submission  
JOURNAL Submitted (25-FEB-2004) Genoscope - Centre National de Sequencage ;  
BP 191 91006 Evry cedex - FRANCE (E-mail : sequef@genoscope.cns.fr  
- Web : www.genoscope.cns.fr)  
FEATURES  
source  
1..746  
/organism="Medicago truncatula"  
/mol\_type="genomic DNA"  
/cultivar="Jemalong A17"  
/db\_xref="taxon:3880"  
/clone\_1lb="MTE1"  
/note="Vector: pindigobAC ; Site\_1: EcoRI ; Site\_2: EcoRI  
; Debelle F. and Chalhoub B.-Genoscope sequence ID :  
mte1-21p8fml"

ORIGIN  
Query Match 24.2%; Score 33.6; DB 9; Length 746;  
Best Local Similarity 61.4%; Pred. No. 23;  
Matches 54; Conservative 0; Mismatches 34; Indels 0; Gaps 0;  
QY 50 TCAGAAAGAGGTGCTATTGCTCGTTATTATCTGTTATCCCAAGACCATAT 109  
|||||  
DB 517 TGATTAAAGAGGTGATGATGATGATGATGATGATGATGATGATGATGAT 576  
QY 110 CAAGCTAGACTGTTCTTATTGTTTACA 137  
|||||  
DB 577 CATGATATATCTATATCCAAATGAA 604

Search completed: August 4, 2005, 01:47:13.  
Job time : 7078 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 3, 2005, 20:47:40 ; Search time 276 Seconds  
(without alignment)  
2981.319 Million cell updates/sec

Title: US-10-070-882a-2

Perfect score: 139  
Sequence: 1 gtagctctgctgacgaact.....ctgtcttactgttaacaca 139

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_16Dec04:\*  
1: geneseqn1808:\*  
2: geneseqn1908:\*  
3: geneseqn2008:\*  
4: geneseqn2018:\*  
5: geneseqn2018:\*  
6: geneseqn2002as:\*  
7: geneseqn2002bs:\*  
8: geneseqn2003as:\*  
9: geneseqn2003bs:\*  
10: geneseqn2003cs:\*  
11: geneseqn2003ds:\*  
12: geneseqn2004as:\*  
13: geneseqn2004bs:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysts of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	139	100.0	139	4 AAF82356	Aaf82356 Salmonella
2	139	100.0	2802	12 ADG31136	Adg31136 Salmonella
3	40	28.8	41	6 ABL57265	Ab157265 Escherich
4	37.2	26.8	1385	4 ACA49482	Aca49482 Prokaryot
5	32.6	23.5	266145	10 ADE87477	Ade87477 Fowlpox v
6	31	22.3	772	6 ABQ72598	Abq72598 Human MDD
7	31	22.3	810	6 ABQ72685	Abq72685 Human MDD
8	30.8	22.2	19521	4 AAK81193	Aak81193 Human Imm
9	30.8	22.2	19521	8 ADA41637	Ada41637 Human sec
10	30.8	22.2	19521	10 ADA57769	Ada57769 BAC fragm
11	30.6	22.0	1539	8 ACA02032	Aca02032 C. glutam
12	30.6	22.0	2799	5 AAH67957	Aah67957 C. glutami
13	30.6	22.0	349980	5 AAH68533	Aah68533 C. glutami
14	30.4	21.9	31241	10 AAD63515	Aad63515 Mycoplasma
15	30.4	21.9	31241	10 ACC69145	Acc69145 M. genital
16	30.4	21.9	31241	12 ADN48950	Adn48950 Mycoplasma
17	30.4	21.9	104644	6 ABQ99653	Abq99653 Human MS4
18	30.4	21.9	109559	13 ABD33157	Abd33157 Murine ca
19	30.2	21.7	110000	9 ADB12064_03	Continuation (4 of
20	30.2	21.7	138363	13 ABD32624	Abd32624 Human can

21	30	21.6	622	12 ADJ38283	Adj38283 Plaetid d
22	30	21.6	10711	4 AAK74790	Aak74790 Human imm
23	29.8	21.4	301	4 AAB06526	Aab06526 Human imm
24	29.8	21.4	301	4 AAB93642	Aab93642 Human pro
25	29.8	21.4	301	4 AAB63734	Aab63734 Human pro
26	29.8	21.4	301	4 AAB02707	Aab02707 Prostate
27	29.8	21.4	301	4 AAB84956	Aab84956 Human pro
28	29.8	21.4	301	5 ACA59543	ACA59543 Prostate
29	29.8	21.4	301	6 AB195106	Ab195106 Human P8D
30	29.8	21.4	301	8 ACC95270	Acc95270 Prostate
31	29.8	21.4	301	10 ADB13743	Adb13743 Human pro
32	29.8	21.4	301	10 ADG26159	Adg26159 Human pro
33	29.8	21.4	340	5 ABV07484	Abv07484 Human pro
34	29.8	21.4	409	5 ABV37417	Abv37417 Human pro
35	29.8	21.4	434	7 ADS71710	AdS71710 Human kid
36	29.8	21.4	434	7 ADS71943	AdS71943 Human kid
37	29.8	21.4	492	6 ABL64736	Ab164736 Lung canc
38	29.8	21.4	492	6 ABL65652	Ab165652 Lung canc
39	29.8	21.4	1140	10 ADC92238	Adc92238 E. faeciu
40	29.8	21.4	1448	2 AAZ33468	Aaz33468 Human pro
41	29.8	21.4	2143	2 AAZ42062	Aaz42062 Human end
42	29.8	21.4	2654	6 ABB57323	Abb57323 CDNA enco
43	29.8	21.4	2700	4 ABL28364	Ab128364 Drosophila
44	29.8	21.4	2840	12 ADP64619	Adp64619 DNA of hu
45	29.8	21.4	2962	10 ADB63189	Adb63189 Human CDN

## ALIGNMENTS

RESULT 1	AAF82356	standard; DNA; 139 BP.
ID	AAF82356	standard; DNA; 139 BP.
XX	AAF82356	
AC	AAF82356	
XX	AAF82356	
DT	22-JUN-2001	(first entry)
XX	22-JUN-2001	
DE	Salmonella typhimurium phoP gene promoter.	
XX	Salmonella typhimurium; phoP gene promoter; PphoP; antibacterial;	
KW	antiviral; vaccine; lacZ; pagC; outer membrane porin C; ompC;	
KW	transgenic microorganism; antigen production; antigen delivery;	
KW	infection; ds.	
XX	Salmonella typhimurium.	
OS	Salmonella typhimurium.	
XX	Salmonella typhimurium.	
PN	WO200119974-A2.	
XX	22-MAR-2001.	
XX	06-SEP-2000; 2000WO-GB003402.	
PF	06-SEP-2000; 2000WO-GB003402.	
XX	10-SEP-1999; 99GB-00021275.	
PR	12-JUL-2000; 2000GB-00017000.	
XX	(MINA ) UK SEC FOR DEFENCE.	
PA	Ttbbal1 RW, Bullifent HL;	
PI	Ttbbal1 RW, Bullifent HL;	
XX	WPI; 2001-328017/34.	
DR	WPI; 2001-328017/34.	
XX	New recombinant gut-colonizing microorganism, useful as vaccine	
PT	component, comprises construct containing phoP, pagC or ompC gene	
PT	promoter linked to nucleic acid encoding protein that induces immune	
PT	response against pathogen.	
XX	Claim 1, Fig 6; 33pp; English.	
PS	The present sequence is a DNA fragment comprising the phoP gene promoter.	
XX	CC DNA fragments comprising the Salmonella typhimurium phoP, pagC, and ompC	
CC	gene promoters were integrated into a vector used to transform a	
CC	recombinant gut-colonizing microorganism. The promoter was operably	

CC linked to a nucleic acid encoding a protein that is able to induce a  
CC protective immune response against an organism in a mammal. The construct  
CC is useful for enhancing expression of a desired protein at mucosal  
CC effector sites. It is useful for delivering a variety of antigenic agents  
CC which can be used to induce a protective immune response against a wide  
CC range of pathogens such as *Bacillus anthracis*, *Bordetella pertussis*,  
CC *Schistosoma mansoni*, herpes simplex virus, and *Mycobacterium tuberculosis*.  
CC The three promoters (P(phop), P(pagC) and P(ompC)) are induced at  
CC different stages in the infection process, and hence at different sites  
CC in the body. This approach allows the induction of different immune  
CC responses which provide protection against pathogens which colonise  
CC different host cell compartments. The *Salmonella* vaccine vector system is  
CC ideally suited to the delivery of many vaccine antigens since the vaccine  
CC delivery mechanism accurately mimics the natural disease, entering the  
CC body via the gut

XX  
SQ Sequence 139 BP; 37 A; 33 C; 23 G; 46 T; 0 U; 0 Other;  
Query Match 100.0%; Score 139; DB 4; Length 139;  
Best Local Similarity 100.0%; Pred. No. 6.3e-36;  
Matches 139; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGACTCTGCTGCGAGCACTTAATATATGCTGCTCACCCTCTTTCTTCAGAAAGAG 60  
DB 1 GTGACTCTGCTGCGAGCACTTAATATATGCTGCTCACCCTCTTTCTTCAGAAAGAG 60

QY 61 GTGACTATTGTCGTGTTTATTACTGTTATCCCAAGACCATATCAACGCTAGAC 120  
DB 61 GTGACTATTGTCGTGTTTATTACTGTTATCCCAAGACCATATCAACGCTAGAC 120

QY 121 TGTCTTATTGTTTAACACA 139  
DB 121 TGTCTTATTGTTTAACACA 139

RESULT 2  
ADG31136  
ID ADG31136 standard; DNA; 2802 BP.  
XX  
AC ADG31136;  
XX  
DT 26-FEB-2004 (first entry)  
XX  
XX  
DE *Salmonella typhimurium* PphoP and phoP operon DNA.  
XX  
XX  
KW live attenuated derivative; pathogenic Enterobacteriaceae;  
KW cross protective immunity; antibacterial; immunostimulant; vaccine;  
XX PphoP; promoter; phoP operon; ds; gene.  
OS *Salmonella typhimurium*.  
XX  
PN WO2003096812-A1.  
XX  
PD 27-NOV-2003.  
XX  
PF 15-APR-2003; 2003WO-US011802.  
XX  
PR 15-APR-2002; 2002US-0372616P.  
PR 18-APR-2002; 2002US-0373626P.  
XX  
PA (UNIW ) UNIV WASHINGTON.  
XX  
PI Curtiss R;  
XX  
DR WPI; 2004-042484/04.  
DR P-PSDB; ADG31137, ADG31138, ADG31139.  
XX  
XX  
PT New live attenuated derivative of a pathogenic Enterobacteriaceae  
PT species, useful as a vaccine for inducing cross protective immunity  
PT against infections caused by various Enterobacteriaceae strains or  
PT serotypes.  
XX  
XX Example 17; Fig 28; 133pp; English.

XX  
CC The invention relates to a novel live attenuated derivative of a  
CC pathogenic Enterobacteriaceae species having enhanced ability to induce  
CC cross protective immunity against Enterobacteriaceae. The derivative of  
CC the invention demonstrates antibacterial and immunostimulant activities  
CC and may be useful as a vaccine for inducing a high level immune response  
CC and/or cross protective immune response to protect individuals from  
CC infection from a diversity of species or serotypes of bacterial  
CC pathogens. The current sequence is that of the *Salmonella typhimurium*  
CC PphoP and phoP operon DNA of the invention.

XX  
SQ Sequence 2802 BP; 721 A; 702 C; 724 G; 655 T; 0 U; 0 Other;  
Query Match 100.0%; Score 139; DB 12; Length 2802;  
Best Local Similarity 100.0%; Pred. No. 1.5e-35;  
Matches 139; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGACTCTGCTGCGAGCACTTAATATATGCTGCTCACCCTCTTTCTTCAGAAAGAG 60  
DB 511 GTGACTCTGCTGCGAGCACTTAATATATGCTGCTCACCCTCTTTCTTCAGAAAGAG 570

QY 61 GTGACTATTGTCGTGTTTATTACTGTTATCCCAAGACCATATCAACGCTAGAC 120  
DB 571 GTGACTATTGTCGTGTTTATTACTGTTATCCCAAGACCATATCAACGCTAGAC 630

QY 121 TGTCTTATTGTTTAACACA 139  
DB 631 TGTCTTATTGTTTAACACA 649

RESULT 3  
ABL57265  
ID ABL57265 standard; DNA; 41 BP.  
XX  
XX  
AC ABL57265;  
XX  
DT 09-AUG-2002 (first entry)  
XX  
XX  
DE *Escherichia coli* DNA 5' to phoA coding region.  
XX  
XX  
KW Transcription terminator; trpA; attenuation; vaccine; virulence;  
KW antibacterial; fungicide; antiparasitic; protozoacide; phoA; ds.  
XX  
OS *Escherichia coli*.  
XX  
XX  
FH Key Location/Qualifiers  
FT misc\_feature 1..19  
FT /\*tag= a  
FT /note= "5' single-stranded overhang"  
FT misc\_feature 41  
FT FT /\*tag= b  
FT /note= "5' overhang on complementary strand of 4 bases  
FT with sequence 5'-GATC-3'."

XX  
XX  
XX WO200230457-A2.  
XX  
XX  
XX PD 16-APR-2002.  
XX  
XX PF 11-OCT-2001; 2001WO-US031606.  
XX  
XX PR 12-OCT-2000; 2000US-00689123.  
XX  
XX  
XX (UNIW ) UNIV WASHINGTON.  
XX PA (MEGA-) MEGAN HEALTH INC.  
XX PA (CURT/) CURTIS R.  
XX PA (TING/) TINGE S A.  
XX  
XX  
XX Curtiss R, Tinge SA;  
XX  
XX  
XX WPI; 2002-444150/47.  
XX  
XX  
XX Composition comprising microbe having attenuating mutation that comprises  
XX insertion sequence containing recombinant transcription terminator,  
PT

PT useful as vaccine, and for delivering a desired gene product to  
PT individual.

XX Example 5; Fig 11; 91pp; English.

XX The present sequence is that of DNA located 5' to the phoA coding  
CC sequence of Escherichia coli. The sequence was used to illustrate an  
CC example of the invention relating to deletion of the phoA gene and  
CC replacement with a trpA terminator. The phoA mutation was introduced into  
CC Salmonella typhimurium SL1344 to produce strain MGN-1362. The present  
CC invention is based on the discovery that transcription terminators (TTs)  
CC can be inserted in a bacterial gene in an attenuating strategy which not  
CC only attenuates the virulence of the bacteria, but also serves to  
CC restrict the effect of the attenuating mutation to the target gene or  
CC operon of the bacteria. A claimed vaccine comprises a microorganism  
CC having an attenuating mutation in a chromosomal gene, the mutation  
CC comprising an insertion sequence which contains a recombinant TT, such as  
CC trpA. The microorganism is preferably Salmonella, Shigella or  
CC Escherichia, and the TT is preferably inserted in the phoA gene. The  
CC vaccine may further comprise a recombinant gene encoding a desired gene  
CC product from a virus, bacterium, protozoan, parasite or fungus, or  
CC encodes an autoantigen, gamete-specific antigen or an allergen. The  
CC attenuated microorganism is also used in a claimed method for delivering  
CC a desired gene product to an individual

SQ Sequence 41 BP; 10 A; 13 C; 4 G; 14 T; 0 U; 0 Other;

Query Match 28.8%; Score 40; DB 6; Length 41;

Best Local Similarity 100.0%; Pred. No. 0.0024; Mismatches 0; Indels 0; Gaps 0;

Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

11 TCGACGAACTTAATATATGCTGCTCACCCTCTTTCTT 50

1 TCGACGAACTTAATATATGCTGCTCACCCTCTTTCTT 40

Db

RESULT 4

ACA49482

ID ACA49482 standard; DNA; 1385 BP.

XX ACA49482;

XX 19-JUN-2003 (first entry)

XX Prokaryotic essential gene #31139.

XX Antisense; ds; prokaryotic essential gene; cell proliferation;

XX drug design; gene.

XX Salmoneilla paratyphi.

XX WO200277183-A2.

XX 03-OCT-2002.

XX 21-MAR-2002; 2002WO-US009107.

XX 21-MAR-2001; 2001US-00815242.

XX 06-SEP-2001; 2001US-0094893.

XX 25-OCT-2001; 2001US-0342923P.

XX 08-FEB-2002; 2002US-00072851.

XX 06-MAR-2002; 2002US-0362699P.

XX (ELIT-) ELITRA PHARM INC.

XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;

XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX WPI; 2003-029926/02.

XX P-PSDB; ABU45612.

XX New antisense nucleic acids, useful for identifying proteins or screening

XX for homologous nucleic acids required for cellular proliferation to

PT isolate candidate molecules for rational drug discovery programs.  
XX Claim 14; SEQ ID NO 37352; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway of  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-regulated gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than S. aureus, S. typhimurium,  
CC K. pneumoniae or P. aeruginosa. The present sequence is one of the target  
CC prokaryotic essential genes. Note: The sequence data for this patent did  
CC not form part of the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

SQ Sequence 1385 BP; 353 A; 372 C; 335 G; 321 T; 0 U; 4 Other;

Query Match 26.8%; Score 37.2; DB 8; Length 1385;

Best Local Similarity 92.9%; Pred. No. 0.054; Mismatches 39; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

1 GTGACTCTGTGTCGACGAACTTAATATATGCTCACCCT 42

1344 GTGACTCTGTGTCGACGAACTTAATATATGCTCACCCT 1385

Db

RESULT 5

ADE87477

ID ADE87477 standard; DNA; 266145 BP.

XX ADE87477;

XX 29-JAN-2004 (first entry)

XX Fowlpox virus genome DNA.

XX Fowlpox virus; FPV; virulence; tuberculoostatic; protozoacide; antipylretic;

XX cytolactic; hepatotropic; antibacterial; vaccine; malaria; tuberculoostic;

XX East Coast fever; avipox virus; influenza; hepatitis;

XX human papilloma virus; tumour; leishmaniasis; listeriosis; theileria;

XX gene; ds.

XX Fowlpox virus.

XX WO2003047617-A2.

XX 02-DEC-2002; 2002WO-GB005411.

XX 30-NOV-2001; 2001GB-00028733.

XX 30-NOV-2001; 2001US-0334649P.

PA	(ISIS-) ISIS INNOVATION LTD.
XX	
PT	Laidlaw S, Skinner M, Hill A, Gilbert S, Anderson R;
XX	
DR	WPI; 2003-513700/48.
XX	
PT	Treating and/or preventing e.g. malaria or tuberculosis, or eliciting an
PT	immune response, comprises administering a priming composition and a
PT	boosting composition containing a non-replicating viral vector in either
PT	order.
PS	Claim 30; SEQ ID NO 1; 3032pp; English.
XX	
CC	The invention relates to a fowlpox virus (FPV) genome which has
CC	modifications in one or more wild-type FPV genes. The invention further
CC	relates to a novel method for treating and/or preventing a disease in a
CC	subject comprising administering two compositions, each containing a non-
CC	replicating viral vector. At least one of the compositions comprises a
CC	poxvirus vector derived from a fowlpox virus. The novel compositions have
CC	the following activities: virucide, tuberculostatic, protozoacide,
CC	antipruritic, cytostatic, hepatotropic, and antibacterial. The non-
CC	replicating viral vector is useful in a vaccine for an animal,
CC	particularly a mammal such as a primate, specifically human. The priming
CC	or boosting composition, or the kit is useful for manufacturing a
CC	medicament for treating and/or preventing a disease which is, or results
CC	from, a chronic infection such as malaria, tuberculosis or East Coast
CC	fever, or for eliciting a T-cell immune response in a subject. Non-
CC	cultured CER cells are useful for growing an avipox virus, such as
CC	fowlpox virus. The method or the vaccine may further be used to treat or
CC	prevent influenza, hepatitis, human papilloma virus and other viral
CC	infections, malignancies such as tumours, leishmaniasis, listeriosis, and
CC	chlamydia. This polynucleotide sequence represents the DNA of the fowlpox
XX	virus genome of the invention.
XX	
SQ	Sequence 266145 BP; 92375 A; 41026 C; 40846 G; 91897 T; 0 U; 1 Other;
XX	
Query Match	23.5%; Score 32.6; DB 10; Length 266145;
Best Local Similarity	55.9%; Pred. No. 8;
Matches	62; Conservative 0; Mismatches 49; Indels 0; Gaps 0;
QY	18 ACTTAATAAAGCGCTGCTACCCCTCTTTCTTCAGAAAGGGTACTATTGCTGCT 77
DB	21181 AGTACTATAGATATATATTAAATCTTATCTGACGAGGAGCAAGATCCATGATATG 212404
QY	78 TTATTAACGTGTTATCCCAAGACATATACAGCGCTAGACTGTTCTTA 128
DB	21241 TTATGATGTTCTGTCATTAAGGGTAAATGATCATATCTTAATGTTTTTA 212931
RESULT 6	
ID	ABQ72598/C
AC	ABQ72598 standard; cDNA; 772 BP.
XX	
AC	ABQ72598;
DT	
XX	
DT	03-SEP-2002 (first entry)
XX	
DE	Human MDDT encoding cDNA SEQ ID NO 150.
XX	
XX	Human; MDDT; disease detection and treatment molecule polynucleotide;
KW	proliferative disorder; hepatitis; psoriasis; cancer; AIDS;
KW	autoimmune disorder; inflammatory disorder; allergy; multiple sclerosis;
KW	rheumatoid arthritis; transgenic; gene therapy; antiarteriosclerotic;
KW	hepatotropic; antiinflammatory; antipsoriatic; cytostatic; anti-HIV;
KW	antiallergic; antianaemic; antischmatic; antischlerotic; antitumor;
XX	neuroprotective; antineumatic; antiarthritic; gene; ss.
XX	
OS	Homo sapiens.
XX	
FN	WO200240715-A2.
XX	
PD	23-MAY-2002.

06-SEP-2001; 2001WO-US027628.

05-SEP-2000; 2000US-0229747P.  
05-SEP-2000; 2000US-0229748P.  
05-SEP-2000; 2000US-0229749P.  
05-SEP-2000; 2000US-0229750P.  
05-SEP-2000; 2000US-0229751P.  
05-SEP-2000; 2000US-0230583P.  
06-SEP-2000; 2000US-0230505P.  
06-SEP-2000; 2000US-0230514P.  
06-SEP-2000; 2000US-0230515P.  
06-SEP-2000; 2000US-0230517P.  
06-SEP-2000; 2000US-0230518P.  
06-SEP-2000; 2000US-0230599P.  
06-SEP-2000; 2000US-0230610P.  
06-SEP-2000; 2000US-0230865P.  
06-SEP-2000; 2000US-0230988P.  
06-SEP-2000; 2000US-0230989P.  
07-SEP-2000; 2000US-0230951P.  
07-SEP-2000; 2000US-0231163P.  
07-SEP-2000; 2000US-0231167P.

(INCYTE GENOMICS INC.

Jackson S, Lincoln SE, Altus CM, Dufour GE, Chalup MS, Hillman JL, Jones AL, Yu JY, Wright RJ, Glatzen D, Liu TF, Yap PE, Dahl CR, Momiyama MG, Bradley DL, Rohatgi SD, Harris B, Roseberry AM, Gerstlin EH, Peralta CH, David MH, Panter SR, Flores V, Daffo A, Marwaha R, Chen AJ, Chang SC, Au AP, Iman RR; WPI; 2002-527544/56.  
P-PSDB; ABP51381.

Novel human disease detection and treatment polypeptide, useful in diagnosis, prevention or treatment of cell proliferative disorders e.g. arteriosclerosis, cirrhosis and an autoimmune/inflammatory disorder e.g. AIDS.

Claim 1, Page 390-391, 618pp; English.

The invention relates to an isolated human disease detection and treatment (MDMT) polypeptide (I) selected from a polypeptide having a sequence selected from 254 sequences (ABP51231-ABP5184) given in the specification, a naturally occurring polypeptide comprising a sequence having at least 90% identity to (I) or a biologically active or immunogenic fragment of (I). (I) is useful for screening a compound for effectiveness as an agonist or antagonist, for screening a compound that specifically binds (I) or modulates the activity of (I), and for preparing a polyclonal or monoclonal antibody by hybridoma technology. Nucleic acids (II) (ABQ7449-ABQ72700) encoding (I) are useful for screening a compound for effectiveness in altering expression of a target polynucleotide comprising. Oligonucleotides and antibodies are useful for detecting MDMT in a sample or for assessing toxicity of a test compound, in a diagnostic test for a condition or a disease associated with the expression of MDMT in a biological sample, for detecting (I) in a sample, and for purifying (I) from a sample. A composition comprising (I), an agonist or antagonist is useful for treating a disease or condition associated with decreased or increased expression of functional MDMT. (I) or (II) are useful for diagnosing, treating or preventing disorders associated with aberrant expression of MDMT, where the disorders are selected from a cell proliferative disorder such as arteriosclerosis, cirrhosis, hepatitis, psoriasis, and cancer and an autoimmune/inflammatory disorder such as AIDS, Addison's disease, allergy, anaemia, asthma, arteriosclerosis, gout, multiple sclerosis or rheumatoid arthritis. (II) are useful for creating knockin humanised animals or transgenic animals to model human diseases, in somatic or germ-line gene therapy, to generate a transcribe image of a tissue or cell type, for detecting differences in the chromosomal location due to

CC translocation or inversion among normal, carrier or affected individuals  
CC and as hybridisation probes for mapping naturally occurring genomic  
CC sequences

XX Sequence 772 BP; 252 A; 147 C; 156 G; 217 T; 0 U; 0 Other;

XX Query Match 22.3%; Score 31; DB 6; Length 772;

XX Best Local Similarity 57.9%; Pred. No. 5.1;

XX Matches 55; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

XX Db 45 TTCTCTCAGAAAGAGGCTGCTGTTTGTCTGTTTAACTGTTATCCCAAGCACC 104

XX 549 TTCCATTAATAAATAAATCCCTTACCTTATTTATATACAGTTTGTCCAAAGCGCT 490

XX 105 ATATCAACGCTAGACTGTTCTTATTTGTTAACA 139

XX 489 TTCTGTAATGCTAGAAATCTCTCAGTCAACACA 455

XX Db

XX RESULT 7

XX ABQ72685/c

XX ID ABQ72685 standard; cDNA; 810 BP.

XX AC ABQ72685;

XX DT 03-SEP-2002 (first entry)

XX DE Human MDDT encoding cDNA SEQ ID NO 237.

XX XX Human: MDDT; disease detection and treatment molecule polynucleotide;

XX KM proliferative disorder; hepatitis; psoriasis; cancer; AIDS;

XX KM rheumatoid arthritis; transgenic; gene therapy; antiretroviral;

XX KM hepatocellular carcinoma; antiproliferative; cytotoxic; anti-HIV;

XX KM antiangiogenic; antineoplastic; antitumor; antithrombotic; antitumor;

XX KM neuroprotective; antineoplastic; antitumor; antitumor; gene; ss.

XX OS Homo sapiens.

XX PN MO200240715-A2.

XX PD 23-MAY-2002.

XX PF 06-SEP-2001; 2001WO-US027628.

XX PR 05-SEP-2000; 2000US-0229747P.

XX PR 05-SEP-2000; 2000US-0229748P.

XX PR 05-SEP-2000; 2000US-0229750P.

XX PR 05-SEP-2000; 2000US-0229751P.

XX PR 05-SEP-2000; 2000US-0230583P.

XX PR 06-SEP-2000; 2000US-0230585P.

XX PR 06-SEP-2000; 2000US-0230587P.

XX PR 06-SEP-2000; 2000US-0230589P.

XX PR 06-SEP-2000; 2000US-0230591P.

XX PR 06-SEP-2000; 2000US-0230593P.

XX PR 06-SEP-2000; 2000US-0230595P.

XX PR 06-SEP-2000; 2000US-0230597P.

PI Dahl CR, Momiyama MG, Bradley DL, Rohatgi SD, Harris B;  
PI Rosberry AM, Gershtin EH, Peralta CH, David MH, Panzer SR, Flores V;  
PI Daffo A, Marwaha R, Chen AJ, Chang SC, Au AP, Inman RR;  
XX WPI; 2002-527544/56.  
XX P-PSDB; ABP51469.

XX DR Novel human disease detection and treatment polypeptide, useful in

XX PT diagnosis, prevention or treatment of cell proliferative disorders e.g.

XX PT arteriosclerosis, cirrhosis and an autoimmune/inflammatory disorder e.g.

XX PT AIDS.

XX PS Claim 1, Page 435; 618pp; English.

XX XX The invention relates to an isolated human disease detection and

XX CC treatment (MDDT) polypeptide (I) selected from a polypeptide having a

XX CC sequence selected from 254 sequences (ABP51211-ABP51484) given in the

XX CC specification, a naturally occurring polypeptide comprising a sequence

XX CC having at least 90% identity to (I) or a biologically active or

XX CC immunogenic fragment of (I). (I) is useful for screening a compound for

XX CC effectiveness as an agonist or antagonist, for screening a compound that

XX CC specifically binds (I) or modulates the activity of (I), and for

XX CC preparing a polyclonal or monoclonal antibody by hybridoma technology.

XX CC Nucleic acids (II) (ABQ72449-ABQ72700) encoding (I) are useful for

XX CC screening a compound for effectiveness in altering expression of a target

XX CC polynucleotide comprising. Oligonucleotides and antibodies are useful for

XX CC detecting MDDT in a sample or for assessing toxicity of a test compound,

XX CC in a diagnostic test for a condition or a disease associated with the

XX CC expression of MDDT in a biological sample, for detecting (I) in a sample,

XX CC and for purifying (I) from a sample. A composition comprising (I), an

XX CC agonist or antagonist is useful for treating a disease or condition

XX CC associated with decreased or increased expression of functional MDDT. (I)

XX CC or (II) are useful for diagnosing, treating or preventing disorders

XX CC associated with aberrant expression of MDDT, where the disorders are

XX CC selected from a cell proliferative disorder such as arteriosclerosis,

XX CC cirrhosis, hepatitis, psoriasis, and cancer and an

XX CC autoimmune/inflammatory disorder such as AIDS, Addison's disease,

XX CC allergy, anaemia, asthma, atherosclerosis, gout, multiple sclerosis or

XX CC rheumatoid arthritis. (II) are useful for creating knockin humanised

XX CC animals or transgenic animals to model human diseases, in somatic or

XX CC germline gene therapy, to generate a transcript image of a tissue or cell

XX CC type, for detecting differences in the chromosomal location due to

XX CC translocation or inversion among normal, carrier or affected individuals

XX CC and as hybridisation probes for mapping naturally occurring genomic

XX CC sequences

XX XX Sequence 810 BP; 262 A; 153 C; 165 G; 230 T; 0 U; 0 Other;

XX XX Query Match 22.3%; Score 31; DB 6; Length 810;

XX XX Best Local Similarity 57.9%; Pred. No. 5.2;

XX XX Matches 55; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

XX XX Db 45 TTCTCTCAGAAAGAGGCTGCTGTTTGTCTGTTTAACTGTTATCCCAAGCACC 104

XX 549 TTCCATTAATAAATAAATCCCTTACCTTATTTATATACAGTTTGTCCAAAGCGCT 490

XX 105 ATATCAACGCTAGACTGTTCTTATTTGTTAACA 139

XX 489 TTCTGTAATGCTAGAAATCTCTCAGTCAACACA 455

XX Db

XX RESULT 8

XX AAK81193/c

XX ID AAK81193 standard; DNA; 19521 BP.

XX AC AAK81193;

XX DT 07-NOV-2001 (first entry)

XX DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:36005.

XX KM Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;

XX KM cytostatic; gene therapy; vaccine; metastasis; ds.

```
XX Homo sapiens.
OS
XX WO200157182-A2.
XX
XX 09-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US001354.
XX
XX 31-JAN-2000; 2000US-0179665P.
XX 04-FEB-2000; 2000US-0180628P.
XX 24-FEB-2000; 2000US-0184664P.
XX 02-MAR-2000; 2000US-0186350P.
XX 16-MAR-2000; 2000US-0189874P.
XX 17-MAR-2000; 2000US-0190076P.
XX 18-APR-2000; 2000US-0198123P.
XX 19-MAY-2000; 2000US-020515P.
XX 07-JUN-2000; 2000US-0209467P.
XX 28-JUN-2000; 2000US-0214886P.
XX 30-JUN-2000; 2000US-0215135P.
XX 07-JUL-2000; 2000US-0216647P.
XX 07-JUL-2000; 2000US-0216880P.
XX 11-JUL-2000; 2000US-0217487P.
XX 11-JUL-2000; 2000US-0217496P.
XX 14-JUL-2000; 2000US-0218290P.
XX 26-JUL-2000; 2000US-0220963P.
XX 26-JUL-2000; 2000US-0220964P.
XX 14-AUG-2000; 2000US-0224518P.
XX 14-AUG-2000; 2000US-0224519P.
XX 14-AUG-2000; 2000US-0225213P.
XX 14-AUG-2000; 2000US-0225214P.
XX 14-AUG-2000; 2000US-0225266P.
XX 14-AUG-2000; 2000US-0225267P.
XX 14-AUG-2000; 2000US-0225268P.
XX 14-AUG-2000; 2000US-0225270P.
XX 14-AUG-2000; 2000US-022547P.
XX 14-AUG-2000; 2000US-0225757P.
XX 14-AUG-2000; 2000US-0225758P.
XX 14-AUG-2000; 2000US-0225759P.
XX 18-AUG-2000; 2000US-0226279P.
XX 22-AUG-2000; 2000US-0226681P.
XX 22-AUG-2000; 2000US-0226688P.
XX 22-AUG-2000; 2000US-0227182P.
XX 23-AUG-2000; 2000US-0227009P.
XX 30-AUG-2000; 2000US-0228924P.
XX 01-SEP-2000; 2000US-0229287P.
XX 01-SEP-2000; 2000US-0229343P.
XX 01-SEP-2000; 2000US-0229344P.
XX 01-SEP-2000; 2000US-0229345P.
XX 05-SEP-2000; 2000US-0229509P.
XX 05-SEP-2000; 2000US-0229513P.
XX 06-SEP-2000; 2000US-0230437P.
XX 06-SEP-2000; 2000US-0230438P.
XX 08-SEP-2000; 2000US-0231242P.
XX 08-SEP-2000; 2000US-0231243P.
XX 08-SEP-2000; 2000US-0231244P.
XX 08-SEP-2000; 2000US-0231413P.
XX 08-SEP-2000; 2000US-0231414P.
XX 08-SEP-2000; 2000US-0232080P.
XX 12-SEP-2000; 2000US-0231968P.
XX 14-SEP-2000; 2000US-0232397P.
XX 14-SEP-2000; 2000US-0232398P.
XX 14-SEP-2000; 2000US-0232399P.
XX 14-SEP-2000; 2000US-0232400P.
XX 14-SEP-2000; 2000US-0232401P.
XX 14-SEP-2000; 2000US-0233063P.
XX 14-SEP-2000; 2000US-0233064P.
XX 14-SEP-2000; 2000US-0233065P.
XX 21-SEP-2000; 2000US-0234223P.
XX 21-SEP-2000; 2000US-0234274P.
XX 25-SEP-2000; 2000US-0234997P.
XX 25-SEP-2000; 2000US-0234998P.
XX
XX 26-SEP-2000; 2000US-0235848P.
XX 27-SEP-2000; 2000US-0235834P.
XX 27-SEP-2000; 2000US-0235836P.
XX 29-SEP-2000; 2000US-0236327P.
XX 29-SEP-2000; 2000US-0236367P.
XX 29-SEP-2000; 2000US-0236368P.
XX 29-SEP-2000; 2000US-0236369P.
XX 29-SEP-2000; 2000US-0236370P.
XX 02-OCT-2000; 2000US-0236802P.
XX 02-OCT-2000; 2000US-0237037P.
XX 02-OCT-2000; 2000US-0237038P.
XX 02-OCT-2000; 2000US-0237039P.
XX 02-OCT-2000; 2000US-0237040P.
XX 13-OCT-2000; 2000US-0239935P.
XX 13-OCT-2000; 2000US-0239937P.
XX 20-OCT-2000; 2000US-0240960P.
XX 20-OCT-2000; 2000US-0241221P.
XX 20-OCT-2000; 2000US-0241222P.
XX 20-OCT-2000; 2000US-0241785P.
XX 20-OCT-2000; 2000US-0241786P.
XX 20-OCT-2000; 2000US-0241787P.
XX 20-OCT-2000; 2000US-0241808P.
XX 20-OCT-2000; 2000US-0241809P.
XX 20-OCT-2000; 2000US-0241826P.
XX 01-NOV-2000; 2000US-0244617P.
XX 08-NOV-2000; 2000US-0246474P.
XX 08-NOV-2000; 2000US-0246475P.
XX 08-NOV-2000; 2000US-0246476P.
XX 08-NOV-2000; 2000US-0246477P.
XX 08-NOV-2000; 2000US-0246478P.
XX 08-NOV-2000; 2000US-0246522P.
XX 08-NOV-2000; 2000US-0246524P.
XX 08-NOV-2000; 2000US-0246525P.
XX 08-NOV-2000; 2000US-0246526P.
XX 08-NOV-2000; 2000US-0246527P.
XX 08-NOV-2000; 2000US-0246528P.
XX 08-NOV-2000; 2000US-0246532P.
XX 08-NOV-2000; 2000US-0246609P.
XX 08-NOV-2000; 2000US-0246610P.
XX 08-NOV-2000; 2000US-0246611P.
XX 08-NOV-2000; 2000US-0246613P.
XX 17-NOV-2000; 2000US-0249207P.
XX 17-NOV-2000; 2000US-0249208P.
XX 17-NOV-2000; 2000US-0249209P.
XX 17-NOV-2000; 2000US-0249210P.
XX 17-NOV-2000; 2000US-0249211P.
XX 17-NOV-2000; 2000US-0249212P.
XX 17-NOV-2000; 2000US-0249213P.
XX 17-NOV-2000; 2000US-0249214P.
XX 17-NOV-2000; 2000US-0249215P.
XX 17-NOV-2000; 2000US-0249216P.
XX 17-NOV-2000; 2000US-0249217P.
XX 17-NOV-2000; 2000US-0249218P.
XX 17-NOV-2000; 2000US-0249244P.
XX 17-NOV-2000; 2000US-0249245P.
XX 17-NOV-2000; 2000US-0249264P.
XX 17-NOV-2000; 2000US-0249265P.
XX 17-NOV-2000; 2000US-0249297P.
XX 17-NOV-2000; 2000US-0249299P.
XX 17-NOV-2000; 2000US-0249300P.
XX 01-DEC-2000; 2000US-0250160P.
XX 01-DEC-2000; 2000US-0250391P.
XX 05-DEC-2000; 2000US-0251030P.
XX 05-DEC-2000; 2000US-0251988P.
XX 05-DEC-2000; 2000US-0256719P.
XX 06-DEC-2000; 2000US-0251479P.
XX 08-DEC-2000; 2000US-0251856P.
XX 08-DEC-2000; 2000US-0251868P.
XX 08-DEC-2000; 2000US-0251869P.
XX 08-DEC-2000; 2000US-0251989P.
XX 08-DEC-2000; 2000US-0251990P.
XX 11-DEC-2000; 2000US-0254097P.
XX 05-JAN-2001; 2001US-0259678P.
XX
```

PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Rosen CA, Barash SC, Ruben SM;  
 XX  
 DR WPI, 2001-483426/52.  
 XX  
 PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
 PT useful for preventing, diagnosing and/or treating cancers and metastasis.  
 XX  
 PS Disclosure; SEQ ID NO 36005; 3071bp + Sequence Listing; English.  
 XX  
 CC AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)  
 CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic  
 CC activity, and can be used in gene therapy and vaccine production. (I)  
 CC proteins and polynucleotides may be used in the prevention, diagnosis and  
 CC treatment of diseases associated with inappropriate (I) expression. For  
 CC example, they may be used to treat disorders associated with decreased  
 CC expression by rectifying mutations or deletions in a patient's genome  
 CC that affect the activity of (I) by expressing inactive proteins or to  
 CC supplement the patient's own production of (I). Additionally, (I)  
 CC polynucleotides may be used to produce the secreted (I), by inserting the  
 CC nucleic acids into a host cell and culturing the cell to express the  
 CC protein. (I) proteins and polynucleotides may be used to prevent,  
 CC diagnose and treat immune/hematopoietic-related diseases, especially  
 CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703  
 CC to AAK67694 represent human immune/hematopoietic antigen genomic  
 CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169  
 CC represent sequences used in the exemplification of the present invention  
 XX  
 SQ Sequence 19521 BP; 5672 A; 3465 C; 3902 G; 6482 T; 0 U; 0 Other;  
 XX  
 Query Match 22.2%; Score 30.8; DB 4; Length 19521;  
 Best Local Similarity 63.5%; Pred. No. 15;  
 Matches 47; Conservative 0; Mismatches 27; Indels 0; Gaps 0;  
 XX  
 QY 1 GTGACTCTGTCGACGAACTTAATAATGCGCTGACCCCTTTCTTCAGAAAGAG 60  
 DB 1311 GTTACTCTGGGCAACTTACATTAACCTCTGTGCTGACCTTGTGACCTGAGAAAGG 1252  
 QY 61 GTGACTATTGTCT 74  
 DB 1251 CTACACATTCCTAT 1238  
 XX  
 RESULT 9  
 ADA1637/c  
 ID ADA1637 standard; DNA; 19521 BP.  
 XX  
 AC ADA1637;  
 XX  
 DT 20-NOV-2003 (first entry)  
 XX  
 DE Human secreted protein related DNA.  
 XX  
 KW Human; secreted protein; cancer; hyperproliferative disorder;  
 KW rheumatoid arthritis; autoimmune disorder; hematopoietic disorder;  
 KW anaemia; allergic reaction; asthma; cardiovascular disorder;  
 KW wound healing; cytostatic; immunosuppressive; neutropenic; neuroprotective;  
 KW antiviral; anti-allergic; hepatotropic; antidiabetic; anti-inflammatory;  
 KW vulnery; cardiac; gene therapy; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN MO2002102993-A2.  
 XX  
 PD 27-DEC-2002.  
 XX  
 PF 19-MAR-2002; 2002MO-US008123.  
 XX  
 PR 21-MAR-2001; 2001US-0277340P.  
 PR 19-JUL-2001; 2001US-0306171P.  
 PR 13-NOV-2001; 2001US-0331287P.  
 XX

PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Rosen CA, Ruben SM;  
 XX  
 DR WPI, 2003-175238/17.  
 XX  
 PT New human secreted proteins and nucleic acid molecules, useful for  
 PT preparing a diagnostic or pharmaceutical composition for diagnosing,  
 PT preventing or treating cancer or other hyperproliferative disorder,  
 PT asthma, allergies or AIDS.  
 XX  
 PS Disclosure; SEQ ID NO 2020; 3205bp; English.  
 XX  
 CC The invention relates to novel genes ADA39629-ADA40565 and proteins  
 CC ADA40566-ADA41501 for human secreted proteins, useful for preventing,  
 CC treating or ameliorating medical conditions e.g. by protein or gene  
 CC therapy. The polypeptides, nucleic acid molecules, antibodies or their  
 CC fragments, and agonists or antagonists that bind to the polypeptide are  
 CC useful for preparing a diagnostic or pharmaceutical composition for  
 CC diagnosing or treating cancer or other hyperproliferative disorder. The  
 CC polypeptides and nucleic acid molecules are also useful for detecting,  
 CC preventing, diagnosing, prognosticating, treating or ameliorating cancer  
 CC or other hyperproliferative disorders including neoplasms, autoimmune  
 CC disorders (e.g. diabetes, rheumatoid arthritis, systemic lupus  
 CC erythematosus, multiple sclerosis, autoimmune thyroiditis or haemolytic  
 CC anaemia), hematopoietic or hematological disorders (e.g. anaemia,  
 CC thrombocytopenia), allergic reactions including asthma or eczema,  
 CC inflammatory disorders (e.g. ischaemia-reperfusion injury, inflammatory  
 CC bowel disease or Crohn's disease), neurodegenerative disorders (e.g.  
 CC Alzheimer's disease or Parkinson's disease), cardiovascular disorders  
 CC (e.g. atherosclerosis, myocarditis), infectious diseases (bacterial,  
 CC fungal or viral infections including HIV/AIDS), or wound healing and  
 CC disorders of epithelial cell proliferation. The nucleic acids are also  
 CC useful for chromosome identification, radiation hybrid mapping or long-  
 CC range restriction mapping, as molecular weight markers, or as  
 CC hybridization or diagnostic probes. The polypeptides and antibodies are  
 CC useful for providing immunological probes for differential identification  
 CC of the tissues immunohistochemistry assays. The sequences given in  
 CC ADA40803-ADA41665 represent DNA sequences related to human secreted  
 CC proteins. Note: The sequence data for this patent did not form part of  
 CC the printed specification, but was obtained in electronic format directly  
 CC from WIPO at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://ftp.wipo.int/pub/published_pct_sequences).  
 XX  
 SQ Sequence 19521 BP; 5672 A; 3465 C; 3902 G; 6482 T; 0 U; 0 Other;  
 XX  
 Query Match 22.2%; Score 30.8; DB 8; Length 19521;  
 Best Local Similarity 63.5%; Pred. No. 15;  
 Matches 47; Conservative 0; Mismatches 27; Indels 0; Gaps 0;  
 XX  
 QY 1 GTGACTCTGTCGACGAACTTAATAATGCGCTGACCCCTTTCTTCAGAAAGAG 60  
 DB 1311 GTTACTCTGGGCAACTTACATTAACCTCTGTGCTGACCTTGTGACCTGAGAAAGG 1252  
 QY 61 GTGACTATTGTCT 74  
 DB 1251 CTACACATTCCTAT 1238  
 XX  
 RESULT 10  
 ADA57769/c  
 ID ADA57769 standard; DNA; 19521 BP.  
 XX  
 AC ADA57769;  
 XX  
 DT 20-NOV-2003 (first entry)  
 XX  
 DE BAC fragment containing human secreted protein gene #505.  
 XX  
 KW Immunosuppressive; anti-inflammatory; antiasthmatic; antiallergic;  
 KW cytotoxic; cerebroprotective; neuroprotective; neutropenic;  
 KW cardiovascular; antiarteriosclerotic; gene therapy;  
 KW human secreted protein; immune disorder; inflammation;  
 KW respiratory disorder; cancer; CNS disorder; neurodegenerative disorders;



KW Corynebacterium; amino acid synthesis; vitamin; saccharide;  
KM organic acid synthesis; ds.  
XX  
XX Corynebacterium glutamicum.  
OS  
XX EPI108790-A2.  
PN  
XX 20-JUN-2001.  
PD  
XX 18-DEC-2000; 2000EP-00127688.  
PF  
XX 16-DEC-1999; 99JP-00377484.  
PR 07-APR-2000; 2000JP-00159162.  
PR 03-AUG-2000; 2000JP-00280988.  
XX  
XX (KYOW ) KYOWA HAKKO KOGYO KK.  
PA  
XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;  
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;  
PI  
XX WPI; 2001-376931/40.  
DR P-PSDB; AAG92738.  
XX  
XX Novel polynucleotides derived from Corynebacterium bacteria, for identifying  
PT mutation point of a gene, measuring expression of a gene, analyzing  
PT expression profile or pattern of a gene and identifying homologous gene.  
PS  
XX Claim 8; SEQ ID NO 2392; 246bp + Sequence Listing; English.  
XX  
XX The present invention provides a number of nucleotide and protein  
CC sequences from the Corynebacterium bacterium Corynebacterium glutamicum. These  
CC are useful for identifying the mutation point of a gene derived from a  
CC mutant of Corynebacterium bacterium, measuring expression amount and analysing  
CC the expression profile or expression pattern of a gene derived from  
CC Corynebacterium bacterium, and identifying a homologue of a gene derived from  
CC Corynebacterium bacterium. Corynebacterium bacteria are useful for producing amino  
CC acids, nucleic acids, vitamins, saccharides and organic acids,  
CC particularly L-lysine. The present sequence is a nucleic acid described  
CC in the exemplification of the invention. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from the European Patent Office  
XX  
SQ Sequence 2799 BP; 715 A; 731 C; 728 G; 625 T; 0 U; 0 Other;  
XX  
XX Query Match 22.0%; Score 30.6; DB 5; Length 2799;  
XX Best Local Similarity 53.8%; Pred. No. 10;  
XX Matches 63; Conservative 0; Mismatches 54; Indels 0; Gaps 0;  
XX  
QY 17 AACTTAATATATGCTGCTCCTCACCCTCTTTCTTCAGAAAGAGGCTATTTGCTGG 76  
DB 271 AACGTCATTGATGGCAGCAGGAGCATGCTTCATGACAGAAACTCAACCATTTGTTGGG 330  
QY 77 TTTATTAAGCTGTTATCCCAAGCAGCATTAATCAAGCTAGCTTTCTTATGTT 133  
DB 331 ATCATGACAGTGTCTGTTTACACCAATTCACCGATCCACGTAATGATCGGT 387  
XX  
XX RESULT 13  
XX AAH68533/c  
XX ID AAH68533 standard; DNA; 349980 BP.  
XX  
XX AAH68533;  
XX  
XX 26-SEP-2001 (first entry)  
XX  
XX C glutamicum coding sequence fragment SEQ ID NO: 7068.  
XX  
XX Corynebacterium; amino acid synthesis; vitamin; saccharide;  
KM organic acid synthesis; ds.  
XX  
XX Corynebacterium glutamicum.  
OS  
XX EPI108790-A2.  
PN

XX  
PD 20-JUN-2001.  
XX  
XX 18-DEC-2000; 2000EP-00127688.  
PF  
XX 16-DEC-1999; 99JP-00377484.  
PR 07-APR-2000; 2000JP-00159162.  
PR 03-AUG-2000; 2000JP-00280988.  
XX  
XX (KYOW ) KYOWA HAKKO KOGYO KK.  
PA  
XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;  
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;  
PI  
XX WPI; 2001-376931/40.  
DR  
XX  
XX Novel polynucleotides derived from Corynebacterium bacteria, for identifying  
PT mutation point of a gene, measuring expression of a gene, analyzing  
PT expression profile or pattern of a gene and identifying homologous gene.  
PS  
XX Disclosure; SEQ ID NO 7068; 246bp + Sequence Listing; English.  
XX  
XX The present invention provides a number of nucleotide and protein  
CC sequences from the Corynebacterium bacterium Corynebacterium glutamicum. These  
CC are useful for identifying the mutation point of a gene derived from a  
CC mutant of Corynebacterium bacterium, measuring expression amount and analysing  
CC the expression profile or expression pattern of a gene derived from  
CC Corynebacterium bacterium, and identifying a homologue of a gene derived from  
CC Corynebacterium bacterium. Corynebacterium bacteria are useful for producing amino  
CC acids, nucleic acids, vitamins, saccharides and organic acids,  
CC particularly L-lysine. The present sequence is a nucleic acid described  
CC in the exemplification of the invention. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from the European Patent Office  
XX  
SQ Sequence 349980 BP; 81250 A; 97718 C; 90621 G; 80391 T; 0 U; 0 Other;  
XX  
XX Query Match 22.0%; Score 30.6; DB 5; Length 349980;  
XX Best Local Similarity 53.8%; Pred. No. 39;  
XX Matches 63; Conservative 0; Mismatches 54; Indels 0; Gaps 0;  
XX  
QY 17 AACTTAATATATGCTGCTCCTCACCCTCTTTCTTCAGAAAGAGGCTATTTGCTGG 76  
DB 200056 AACGTCATTGATGGCAGCAGGAGCATGCTTCATGACAGAAACTCAACCATTTGTTGGG 199997  
QY 77 TTTATTAAGCTGTTATCCCAAGCAGCATTAATCAAGCTAGCTTTCTTATGTT 133  
DB 199996 ATCATGACAGTGTCTGTTTACACCAATTCACCGATCCACGTAATGATCGGT 199940  
XX  
XX RESULT 14  
XX AAD63515/c  
XX ID AAD63515 standard; DNA; 31241 BP.  
XX  
XX AAD63515;  
XX  
XX 12-FEB-2004 (first entry)  
XX  
XX Mycoplasma genitalium gene involved in transport and binding.  
XX  
XX Mycoplasma genitalium genome; bioreactor; bioremediation;  
KM therapeutic biomolecule; energy conversion system; processing system;  
KM anabolic; catabolic system; biological film; cosmetic application;  
KM coating; ds.  
XX  
XX Mycoplasma genitalium.  
OS  
XX US2003138777-A1.  
XX  
XX 24-JUL-2003.  
XX  
XX 20-SEP-2001; 2001US-00960858.  
XX  
XX

```
PR 20-SEP-2001; 2001US-00960858.
XX
XX (EVAN/) EVANS G A.
XX
XX Evans GA;
XX
XX WPI; 2003-851721/79.
XX
XX Basic genetic operating system for an autonomous prototrophic
PT nanomachine, comprises a nanomachine genome encoding a minimal gene set.
XX
XX Example 1; Page 127-141; 170pp; English.
XX
XX The invention relates to a basic genetic operating system which comprises
CC a nanomachine genome encoding a minimal gene set for viability. The basic
CC genetic operating system is used for an autonomous prototrophic
CC nanomachine or autonomous auxotrophic nanomachine. The nanomachine is
CC used as a bioreactor, for bioremediation, for production of a therapeutic
CC biomolecule or as a diagnostic reagent, for production of a diagnostic
CC indicator or as a diagnostic reagent, as a delivery system, as an
CC artificial tissue or organ system, an energy conversion system, as a
CC processing system, as an anabolic or catabolic system, for production of
CC biological films or coatings that may respond to the environment and for
CC cosmetic applications including pharmaceuticals. The present sequence is
CC Mycoplasma genitalium nanomachine gene
SQ Sequence 31241 BP; 10733 A; 4570 C; 5361 G; 10577 T; 0 U; 0 Other;
Query Match 21.9%; Score 30.4; DB 10; Length 31241;
Best Local Similarity 55.8%; Pred. No. 23;
Matches 58; Conservative 0; Mismatches 46; Indels 0; Gaps 0;
QY 34 CCTCACCCCTCTTTCTTCAGAAAGAGGCTACTTGTCTGTTATTACTGTTATC 93
DB 25866 CCTCAGCCCAATTTTATTACCAAGTGGAGATTACCATCTGTTTGTCTTCTGTC 25807
QY 94 CCCAAGCACCATAATCAACGCTAGACTGTTCTTATTGTTACA 137
DB 25806 ATTAAGAAGCGCTATCAATCTATACCCCTTTTGTCTAATA 25763
RESULT 15
ACC69145/C
ID ACC69145 standard; DNA; 31241 BP.
XX
XX ACC69145;
XX
XX 10-JUL-2003 (first entry)
XX
XX M. genitalium transport and binding gene cassette DNA SEQ ID NO:13.
XX
XX Mycoplasma genitalium; gene cassette; replication; transcription;
XX translation; metabolism; basic genetic operating system; gene therapy;
XX autonomous prototrophic nanomachine; auxotrophic nanomachine;
XX nanomachine; bioreactor; bioremediation; therapeutic; delivery system;
XX artificial tissue; artificial organ system; energy conversion system;
XX processing system; anabolic system; catabolic system; biological film;
XX biological coating; cosmetic; gene; ds.
XX
XX Mycoplasma genitalium.
XX
XX WO2003025145-A2.
XX
XX 27-MAR-2003.
XX
XX 18-SEP-2002; 2002WO-US029811.
XX
XX 20-SEP-2001; 2001US-00960870.
XX
XX (EGEA-) EGEA BIOSCIENCES INC.
XX
XX Evans GA;
XX
```

```
DR WPI; 2003-354602/33.
XX
XX New basic genetic operating system for autonomous prototrophic or
PT auxotrophic nanomachine, useful for therapeutic, diagnostic or industrial
PT purposes, comprises a nanomachine genome encoding a gene set for
PT viability or replication.
XX
XX Example 1; Page 231-240; 250pp; English.
XX
XX The present invention describes a basic genetic operating system for an
CC autonomous prototrophic or auxotrophic nanomachine comprising a
CC nanomachine genome encoding a minimal gene set sufficient for viability
CC or replication, optionally in the presence of an auxotrophic molecule.
CC Also described is an autonomous prototrophic or auxotrophic nanomachine
CC comprising a basic genetic operating system for autonomous prototrophic
CC or auxotrophic viability or replication, optionally in the presence of an
CC auxotrophic molecule, and a particle envelope. The nanomachines can be
CC used in gene therapy. The basic genetic operating system or nanomachine
CC is useful in therapeutic, diagnostic and industrial applications, e.g. as
CC a bioreactor, for bioremediation, for the production of a therapeutic
CC biomolecule or as a therapeutic reagent, for the production of a
CC diagnostic indicator or reagent, as a delivery system, as an artificial
CC tissue or organ system, as an energy conversion system, as a processing
CC system, as an anabolic or catabolic system, for the production of
CC biological films or coatings, and for cosmetic applications. The present
CC sequence represents a Mycoplasma genitalium gene cassette nucleotide
CC sequence, which is used in an example from the present invention for the
CC design and synthesis of a basic genetic operation system for a
CC replication competent nanomachine
SQ Sequence 31241 BP; 10733 A; 4570 C; 5361 G; 10577 T; 0 U; 0 Other;
Query Match 21.9%; Score 30.4; DB 10; Length 31241;
Best Local Similarity 55.8%; Pred. No. 23;
Matches 58; Conservative 0; Mismatches 46; Indels 0; Gaps 0;
QY 34 CCTCACCCCTCTTTCTTCAGAAAGAGGCTACTTGTCTGTTATTACTGTTATC 93
DB 25866 CCTCAGCCCAATTTTATTACCAAGTGGAGATTACCATCTGTTTGTCTTCTGTC 25807
QY 94 CCCAAGCACCATAATCAACGCTAGACTGTTCTTATTGTTACA 137
DB 25806 ATTAAGAAGCGCTATCAATCTATACCCCTTTTGTCTAATA 25763
Search completed: August 3, 2005, 21:57:32
Job time : 282 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 3, 2005, 21:54:05 ; Search time 358 Seconds  
(without alignments)  
2513.010 Million cell updates/sec

Title: US-10-070-882a-2

Perfect score: 139  
Sequence: 1 gtgactctgctgacgaact.....ctgtctctatgttaacaca 139

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 7287783 segs, 3236178273 residues

Total number of hits satisfying chosen parameters: 14575566

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications NA.\*

1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq.\*  
10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq.\*  
12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*  
13: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq.\*  
14: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq.\*  
15: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq.\*  
16: /cgn2\_6/ptodata/2/pubpna/US10D\_PUBCOMB.seq.\*  
17: /cgn2\_6/ptodata/2/pubpna/US10E\_PUBCOMB.seq.\*  
18: /cgn2\_6/ptodata/2/pubpna/US10F\_PUBCOMB.seq.\*  
19: /cgn2\_6/ptodata/2/pubpna/US10G\_PUBCOMB.seq.\*  
20: /cgn2\_6/ptodata/2/pubpna/US10H\_PUBCOMB.seq.\*  
21: /cgn2\_6/ptodata/2/pubpna/US10I\_PUBCOMB.seq.\*  
22: /cgn2\_6/ptodata/2/pubpna/US10I\_NEW\_PUB.seq.\*  
23: /cgn2\_6/ptodata/2/pubpna/US11A\_PUBCOMB.seq.\*  
24: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq.\*  
25: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*  
26: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	37.2	26.8	1385	17	US-10-282-122A-37352
2	33	23.7	100596	22	US-10-737-082-92
3	33	23.7	100596	22	US-10-765-790-92
4	32.6	23.5	266145	21	US-10-856-118-1
5	31	22.3	772	19	US-10-363-829-150
6	31	22.3	810	19	US-10-363-829-237
7	31	22.3	2337	21	US-10-486-706-223

8	30.6	22.0	2799	9	US-09-738-626-2992	Sequence 2992, App
9	30.6	22.0	3309400	9	US-09-738-626-1	Sequence 1, Appl
10	30.4	21.9	15549	22	US-10-737-082-4	Sequence 4, Appl
11	30.4	21.9	15649	22	US-10-765-790-4	Sequence 4, Appl
12	30.4	21.9	31241	10	US-09-960-870-13	Sequence 13, Appl
13	30.4	21.9	31241	10	US-09-960-858-13	Sequence 13, Appl
14	30.4	21.9	31241	18	US-10-251-668-13	Sequence 13, Appl
15	30.4	21.9	104644	19	US-10-433-287-79	Sequence 79, Appl
16	30.4	21.9	109559	19	US-10-322-261-137	Sequence 137, Appl
17	30.2	21.7	130349	21	US-10-741-600-17619	Sequence 17619, A
18	30.2	21.7	138363	19	US-10-367-094-117	Sequence 117, App
19	30.2	21.6	458	18	US-10-424-599-105245	Sequence 105245,
20	30	21.6	622	19	US-10-600-070-201	Sequence 201, App
21	29.8	21.4	301	9	US-09-759-143-293	Sequence 293, App
22	29.8	21.4	301	9	US-09-780-665-293	Sequence 293, App
23	29.8	21.4	301	9	US-09-822-827-293	Sequence 293, App
24	29.8	21.4	301	9	US-09-232-880-293	Sequence 293, App
25	29.8	21.4	301	9	US-09-895-793-293	Sequence 293, App
26	29.8	21.4	301	9	US-09-895-814-293	Sequence 293, App
27	29.8	21.4	301	13	US-10-012-896-293	Sequence 293, App
28	29.8	21.4	301	14	US-10-010-940-293	Sequence 293, App
29	29.8	21.4	301	16	US-10-144-678A-293	Sequence 293, App
30	29.8	21.4	301	16	US-10-294-025-293	Sequence 293, App
31	29.8	21.4	409	20	US-10-357-930-7475	Sequence 7475, App
32	29.8	21.4	409	20	US-10-357-930-37436	Sequence 37436, A
33	29.8	21.4	434	15	US-10-102-524-107	Sequence 307, App
34	29.8	21.4	434	15	US-10-102-524-540	Sequence 540, App
35	29.8	21.4	444	13	US-10-027-632-81146	Sequence 81146, A
36	29.8	21.4	444	17	US-10-027-632-81146	Sequence 81146, A
37	29.8	21.4	463	13	US-10-027-632-109769	Sequence 109769,
38	29.8	21.4	463	17	US-10-027-632-109769	Sequence 109769,
39	29.8	21.4	492	9	US-09-954-456-46	Sequence 46, Appl
40	29.8	21.4	492	9	US-09-954-456-962	Sequence 962, App
41	29.8	21.4	492	21	US-10-843-641A-1073	Sequence 3073, App
42	29.8	21.4	492	21	US-10-843-641A-3989	Sequence 3989, App
43	29.8	21.4	617	13	US-10-027-632-229999	Sequence 229999,
44	29.8	21.4	617	17	US-10-027-632-229999	Sequence 229999,
45	29.8	21.4	1400	21	US-10-956-157-7378	Sequence 7378, App

## ALIGNMENTS

RESULT 1  
US-10-282-122A-37352  
; Sequence 37352, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; PRIOR FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347

```

; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 37352
; LENGTH: 1385
; TYPE: DNA
; ORGANISM: Salmonella paratyphi A
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (72)..(72)
; OTHER INFORMATION: n=g, a, t or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (74)..(75)
; OTHER INFORMATION: n=g, a, t or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (158)..(158)
; OTHER INFORMATION: n=g, a, t or c
US-10-282-122A-37352

```

```

Query Match 26.8%; Score 37.2; DB 17; Length 1385;
Best Local Similarity 92.9%; Pred. No. 0.074;
Matches 39; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 1 GGGACCTGGTGGAGCACTTAATATGCGCTGACCCCT 42
DB 1344 GGGACCTGGTGGAGCACTTAATATGCGCTGACCCCT 1385

```

```

RESULT 2
US-10-737-082-92/c
; Sequence 92, Application US/10737082
; Publication No. US20050130170A1
; GENERAL INFORMATION:
; APPLICANT: Bayer Healthcare LLC
; APPLICANT: Beard, Chris
; APPLICANT: Burgess, Chris
; APPLICANT: Gannon, Allison
; APPLICANT: Harvey, Jeanne
; APPLICANT: Lechner, John F.
; APPLICANT: Li, Zheng
; TITLE OF INVENTION: Identification and Verification of Methylation Marker Sequences
; FILE REFERENCE: 1657/2032
; CURRENT APPLICATION NUMBER: US/10/737,082
; PRIOR FILING DATE: 2003-12-16
; PRIOR APPLICATION NUMBER: US 10/737,082
; PRIOR FILING DATE: 2003-12-16
; NUMBER OF SEQ ID NOS: 300
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 92
; LENGTH: 100596
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-737-082-92

```

```

Query Match 23.7%; Score 33; DB 22; Length 100596;
Best Local Similarity 65.8%; Pred. No. 12;
Matches 48; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

```

```

QY 18 ACTTAATATGCGCTGACCCCTGTTTCTTCTTCTGAGAAAGAGGTGACTATTGCTGGT 77

```

```

DB 34052 ATTGAATATGCTGTTTTCCTTCTTTTAAACAATAATGCTAATTAATGACTATT 33993
QY 78 TTATTAAGTGT 90
DB 33992 TTATTAAGTGT 33980

```

```

RESULT 3
US-10-765-790-92/c
; Sequence 92, Application US/10765790
; Publication No. US20050130172A1
; GENERAL INFORMATION:
; APPLICANT: Bayer Healthcare LLC
; APPLICANT: Beard, Chris
; APPLICANT: Burgess, Chris
; APPLICANT: Gannon, Allison
; APPLICANT: Harvey, Jeanne
; APPLICANT: Lechner, John F.
; APPLICANT: Li, Zheng
; TITLE OF INVENTION: Identification and Verification of Methylation Marker Sequences
; FILE REFERENCE: 1657/2035
; CURRENT APPLICATION NUMBER: US/10/765,790
; PRIOR FILING DATE: 2004-01-27
; PRIOR APPLICATION NUMBER: US 10/737,082
; PRIOR FILING DATE: 2003-12-16
; NUMBER OF SEQ ID NOS: 300
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 92
; LENGTH: 100596
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-765-790-92

```

```

Query Match 23.7%; Score 33; DB 22; Length 100596;
Best Local Similarity 65.8%; Pred. No. 12;
Matches 48; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

```

```

QY 18 ACTTAATATGCGCTGACCCCTGTTTCTTCTTCTGAGAAAGAGGTGACTATTGCTGGT 77
DB 34052 ATTGAATATGCTGTTTTCCTTCTTTTAAACAATAATGCTAATTAATGACTATT 33993

```

```

QY 78 TTATTAAGTGT 90
DB 33992 TTATTAAGTGT 33980

```

```

RESULT 4
US-10-856-118-1
; Sequence 1, Application US/10856118
; Publication No. US20050025747A1
; GENERAL INFORMATION:
; APPLICANT: Laidlaw, Stephen
; APPLICANT: Skinner, Mike
; APPLICANT: Hill, Adrian V.S.
; APPLICANT: Gilbert, Sarah C.
; APPLICANT: Anderson, Richard
; TITLE OF INVENTION: Vaccine
; FILE REFERENCE: 3742.1000-000
; CURRENT APPLICATION NUMBER: US/10/856,118
; PRIOR FILING DATE: 2004-05-27
; PRIOR APPLICATION NUMBER: PCT/GB02/005411
; PRIOR FILING DATE: 2002-12-02
; PRIOR APPLICATION NUMBER: GB0128733.3
; PRIOR FILING DATE: 2001-11-30
; PRIOR APPLICATION NUMBER: US 60/334,649
; PRIOR FILING DATE: 2001-11-30
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 266145
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:

```

OTHER INFORMATION: Fowlpox virus genome  
US-10-856-118-1

Query Match 23.5%; Score 32.6; DB 21; Length 266145;  
Best Local Similarity 55.9%; Pred. No. 24;  
Matches 62; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 18 ACTTAATAATGCGCGCTCCACCCCTTTCTTGAGAAAGGGGAGCTATTGTCGTGT 77  
DB 21181 AGTACTATAGATATATTATATCTATATGACGAGGAGACGAAGATCCATGGATGTG 21240

QY 78 TTATTAAGCTTTATCCCAAGACACATTAATCAAGCTAGCTGTCTTA 128  
DB 21241 TTATGATGTTCTGTCTGTCATTAAGGATATGATCATATATATGTTT 21291

RESULT 5  
US-10-363-829-150/c

Sequence 150, Application US/10363829  
Publication No. US2004014231A1

GENERAL INFORMATION:

APPLICANT: Jackson, Stuart E.; Lincoln, Stephen E.;  
APPLICANT: Altus, Christina M.; Dufour, Gerard E.;  
APPLICANT: Chalup, Michael S.; Jackson, Jennifer L.;  
APPLICANT: Jones, Anissa L.; Yu, Jimmy Y.;  
APPLICANT: Wright, Rachel U.; Gietzen, Darryl;  
APPLICANT: Liu, Tommy F.; Yap, Pierre E.;  
APPLICANT: Dahl, Christopher R.; Momiya, Monika G.;  
APPLICANT: Bradley, Diana L.; Rohatgi, Sameer D.;  
APPLICANT: Harris, Bernard; Roseberry Lincoln, Ann M.;  
APPLICANT: Gerstin, Jr., Edward H.; Peralta, Careyna H.;  
APPLICANT: David, Marie H.; Panzer, Scott R.;  
APPLICANT: Flores, Vincent Z.; Daffo, Abel;  
APPLICANT: Marwaha, Rakesh; Chen, Alice J.;  
APPLICANT: Chang, Simon C.; Au, Alan P.;  
APPLICANT: Imman, Rebekah R.

TITLE OF INVENTION: MOLECULES FOR DISEASE DETECTION AND TREATMENT  
FILE REFERENCE: PT-1183 USN

CURRENT APPLICATION NUMBER: US/10/363,829

PRIOR FILING DATE: 2003-03-05

PRIOR APPLICATION NUMBER: PCT/US01/27628

PRIOR FILING DATE: 2001-09-05

PRIOR APPLICATION NUMBER: US 60/229,751

PRIOR FILING DATE: 2000-09-05

PRIOR APPLICATION NUMBER: US 60/229,749

PRIOR FILING DATE: 2000-09-05

PRIOR APPLICATION NUMBER: US 60/229,750

PRIOR FILING DATE: 2000-09-05

PRIOR APPLICATION NUMBER: US 60/229,747

PRIOR FILING DATE: 2000-09-05

PRIOR APPLICATION NUMBER: US 60/229,748

PRIOR FILING DATE: 2000-09-05

PRIOR APPLICATION NUMBER: US 60/230,583

PRIOR FILING DATE: 2000-09-05

PRIOR APPLICATION NUMBER: US 60/230,517

PRIOR FILING DATE: 2000-09-06

PRIOR APPLICATION NUMBER: US 60/230,610

PRIOR FILING DATE: 2000-09-06

PRIOR APPLICATION NUMBER: US 60/230,597

PRIOR FILING DATE: 2000-09-06

Remaining Prior Application data removed - See File Wrapper or PALM.

NUMBER OF SEQ ID NOS: 506

SOFTWARE: PERL Program

SEQ ID NO 150

LENGTH: 772

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: misc feature

OTHER INFORMATION: Incyte ID No: LG:994938.1.2000SEP08

US-10-363-829-150

Query Match 22.3%; Score 31; DB 19; Length 772;

Best Local Similarity 57.9%; Pred. No. 7;  
Matches 55; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 45 TTCTTCGAGAAAGGCGTACTATTGTCGTGTTATTAAGCTTATCCCAAGCAC 104  
DB 549 TTCCAAATTAATAAAGGCTCCATTTAGTTATTTATACAGATTGTCCAAAAGCGCT 490

QY 105 ATAAATCAAGGCTAGCTGTTCTTAATGTTTAACACA 139  
DB 489 TTCTGTATGCTAGAAATCTCTCACTGTCAACACA 455

RESULT 6  
US-10-363-829-237/c

Sequence 237, Application US/10363829  
Publication No. US2004014231A1

GENERAL INFORMATION:

APPLICANT: Jackson, Stuart E.; Lincoln, Stephen E.;  
APPLICANT: Altus, Christina M.; Dufour, Gerard E.;  
APPLICANT: Chalup, Michael S.; Jackson, Jennifer L.;  
APPLICANT: Jones, Anissa L.; Yu, Jimmy Y.;  
APPLICANT: Wright, Rachel U.; Gietzen, Darryl;  
APPLICANT: Liu, Tommy F.; Yap, Pierre E.;  
APPLICANT: Dahl, Christopher R.; Momiya, Monika G.;  
APPLICANT: Bradley, Diana L.; Rohatgi, Sameer D.;  
APPLICANT: Harris, Bernard; Roseberry Lincoln, Ann M.;  
APPLICANT: Gerstin, Jr., Edward H.; Peralta, Careyna H.;  
APPLICANT: David, Marie H.; Panzer, Scott R.;  
APPLICANT: Flores, Vincent Z.; Daffo, Abel;  
APPLICANT: Marwaha, Rakesh; Chen, Alice J.;  
APPLICANT: Chang, Simon C.; Au, Alan P.;  
APPLICANT: Imman, Rebekah R.

TITLE OF INVENTION: MOLECULES FOR DISEASE DETECTION AND TREATMENT  
FILE REFERENCE: PT-1183 USN

CURRENT APPLICATION NUMBER: US/10/363,829

PRIOR FILING DATE: 2003-03-05

PRIOR APPLICATION NUMBER: PCT/US01/27628

PRIOR FILING DATE: 2001-09-05

PRIOR APPLICATION NUMBER: US 60/229,751

PRIOR FILING DATE: 2000-09-05

PRIOR APPLICATION NUMBER: US 60/229,749

PRIOR FILING DATE: 2000-09-05

PRIOR APPLICATION NUMBER: US 60/229,750

PRIOR FILING DATE: 2000-09-05

PRIOR APPLICATION NUMBER: US 60/229,747

PRIOR FILING DATE: 2000-09-05

PRIOR APPLICATION NUMBER: US 60/229,748

PRIOR FILING DATE: 2000-09-05

PRIOR APPLICATION NUMBER: US 60/230,583

PRIOR FILING DATE: 2000-09-05

PRIOR APPLICATION NUMBER: US 60/230,517

PRIOR FILING DATE: 2000-09-06

PRIOR APPLICATION NUMBER: US 60/230,610

PRIOR FILING DATE: 2000-09-06

PRIOR APPLICATION NUMBER: US 60/230,597

PRIOR FILING DATE: 2000-09-06

Remaining Prior Application data removed - See File Wrapper or PALM.

NUMBER OF SEQ ID NOS: 506

SOFTWARE: PERL Program

SEQ ID NO 237

LENGTH: 810

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: misc feature

OTHER INFORMATION: Incyte ID No: LG:814261.1.2000SEP08

US-10-363-829-237

Query Match 22.3%; Score 31; DB 19; Length 810;

Best Local Similarity 57.9%; Pred. No. 7.1;

Matches 55; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 45 TTCTTCGAGAAAGGCGTACTATTGTCGTGTTATTAAGCTTATCCCAAGCAC 104



```
APPLICANT: Burgess, Chris
APPLICANT: Gannon, Allison
APPLICANT: Harvey, Jeanne
APPLICANT: Lechner, John F.
APPLICANT: Li, Zheng
TITLE OF INVENTION: Identification and Verification of Methylation Marker Sequences
FILE REFERENCE: 1657/2032
CURRENT APPLICATION NUMBER: US/10/737,082
CURRENT FILING DATE: 2003-12-16
PRIOR APPLICATION NUMBER: US 10/737,082
NUMBER OF SEQ ID NOS: 300
SOFTWARE: PatentIn version 3.2
SEQ ID NO 4
LENGTH: 15649
TYPE: DNA
ORGANISM: Homo sapiens
US-10-737-082-4
```

Query Match 21.9%; Score 30.4; DB 22; Length 15649;  
Best Local Similarity 53.3%; Pred. No. 40;  
Matches 64; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

```
QY 16 GAACCTAAATATGCGCTGCTCACCCTCTTTCTTCAGAAAGGGTGACTATTGCTG 75
DB 2478 GAACCCAGATATATGCTAGCTGCTTATGATTAAGACTCAGATACATTTTACTG 2419
QY 76 GTTATTAAGTCTTTATCCCAAGCAGCATTAATCAAGCTGACTGTTCTTATGTTAA 135
DB 2418 GGTCAAGTCTGTTTATTAAGAAATTAAGACAGACCCAGCTTCTGATTTCTTAA 2359
```

```
RESULT 11
US-10-765-790-4/c
Sequence 4, Application US/10765790
Publication No. US20050130172A1
GENERAL INFORMATION:
APPLICANT: Bayer Healthcare LLC
APPLICANT: Beard, Chris
APPLICANT: Burgess, Chris
APPLICANT: Gannon, Allison
APPLICANT: Harvey, Jeanne
APPLICANT: Lechner, John F.
APPLICANT: Li, Zheng
TITLE OF INVENTION: Identification and Verification of Methylation Marker Sequences
FILE REFERENCE: 1657/2035
CURRENT APPLICATION NUMBER: US/10/765,790
CURRENT FILING DATE: 2004-01-27
PRIOR APPLICATION NUMBER: US 10/737,082
NUMBER OF SEQ ID NOS: 300
SOFTWARE: PatentIn version 3.2
SEQ ID NO 4
LENGTH: 15649
TYPE: DNA
ORGANISM: Homo sapiens
US-10-765-790-4
```

Query Match 21.9%; Score 30.4; DB 22; Length 15649;  
Best Local Similarity 53.3%; Pred. No. 40;  
Matches 64; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

```
QY 16 GAACCTAAATATGCGCTGCTCACCCTCTTTCTTCAGAAAGGGTGACTATTGCTG 75
DB 2478 GAACCCAGATATATGCTAGCTGCTTATGATTAAGACTCAGATACATTTTACTG 2419
QY 76 GTTATTAAGTCTTTATCCCAAGCAGCATTAATCAAGCTGACTGTTCTTATGTTAA 135
DB 2418 GGTCAAGTCTGTTTATTAAGAAATTAAGACAGACCCAGCTTCTGATTTCTTAA 2359
```

RESULT 12  
US-09-960-870-13/c

```
Sequence 13, Application US/09960870
Publication No. US20030134281A1
GENERAL INFORMATION:
APPLICANT: Evans, Glen
TITLE OF INVENTION: NANOMACHINE COMPOSITIONS AND METHODS OF
FILE REFERENCE: P-EA 4738
CURRENT APPLICATION NUMBER: US/09/960,870
CURRENT FILING DATE: 2001-09-20
NUMBER OF SEQ ID NOS: 19
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 13
LENGTH: 31241
TYPE: DNA
ORGANISM: M. genitalium
US-09-960-870-13
```

Query Match 21.9%; Score 30.4; DB 10; Length 31241;  
Best Local Similarity 55.8%; Pred. No. 53;  
Matches 58; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

```
QY 34 CCTACCCCTCTTTCTTCAGAAAGGGTGACTATTGCTGCTTATTAAGTCTTATC 93
DB 25866 CCTCAGCCATTTTATTAACCAAGTGGTAGTTTACCCTGCTGCTTCTTCTGTC 25807
QY 94 CCCAAGCAGCATTAATCAAGCTGACTGTTCTTATGTTTACA 137
DB 25806 ATTTAAAGAACGCTATCAATCTATACCCCTTTTGTGCTAA 25763
```

```
RESULT 13
US-09-960-858-13/c
Sequence 13, Application US/09960858
Publication No. US20030138777A1
GENERAL INFORMATION:
APPLICANT: Evans, Glen
TITLE OF INVENTION: NANOMACHINE COMPOSITIONS AND METHODS OF
FILE REFERENCE: P-EA 4974
CURRENT APPLICATION NUMBER: US/09/960,858
CURRENT FILING DATE: 2001-09-20
NUMBER OF SEQ ID NOS: 19
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 13
LENGTH: 31241
TYPE: DNA
ORGANISM: M. genitalium
US-09-960-858-13
```

Query Match 21.9%; Score 30.4; DB 10; Length 31241;  
Best Local Similarity 55.8%; Pred. No. 53;  
Matches 58; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

```
QY 34 CCTACCCCTCTTTCTTCAGAAAGGGTGACTATTGCTGCTTATTAAGTCTTATC 93
DB 25866 CCTCAGCCATTTTATTAACCAAGTGGTAGTTTACCCTGCTGCTTCTTCTGTC 25807
QY 94 CCCAAGCAGCATTAATCAAGCTGACTGTTCTTATGTTTACA 137
DB 25806 ATTTAAAGAACGCTATCAATCTATACCCCTTTTGTGCTAA 25763
```

```
RESULT 14
US-10-251-668-13/c
Sequence 13, Application US/10251668
Publication No. US20040063097A1
GENERAL INFORMATION:
APPLICANT: Evans, Glen
TITLE OF INVENTION: NANOMACHINE COMPOSITIONS AND METHODS OF
FILE REFERENCE: P-EA 5441
CURRENT APPLICATION NUMBER: US/10/251,668
CURRENT FILING DATE: 2002-09-20
```

```

PRIORITY APPLICATION NUMBER: US 09/960,607
PRIORITY FILING DATE: 2001-09-20
NUMBER OF SEQ ID NOS: 19
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 13
LENGTH: 31241
TYPE: DNA
ORGANISM: M. genitalium
US-10-251-668-13

Query Match      21.9%; Score 30.4; DB 18; Length 31241;
Best Local Similarity 55.8%; Pred. No. 53;
Matches 58; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

QY      34 CCTACCCCTTTTCTTTCAGAAAGGCTACTATTTGTCGTTTATTAACGTGTTATC 93
Db      25866 CCTCAGCCAAATTTTTTTTAAACCAAGTGCTGAGTTTACCATCGTGTGTTGCTTCTGTGTC 25807
QY      94 CCCAAGACCATATCAACGCTAGCTGTTCTTATGTGTACA 137
Db      25806 ATTAAAGAACGCTATCAATCTATACCCCTTTTGTCTAAAA 25763

RESULT 15
US-10-433-287-79/C
Sequence 79, Application US/10433287
Publication No. US20040137566A1
GENERAL INFORMATION:
APPLICANT: Tedder, Thomas F
APPLICANT: Liang, Yinghua
TITLE OF INVENTION: Identification of Novel Gene Family Members Expressed by
FILE REFERENCE: 180/132 PCT/US
CURRENT APPLICATION NUMBER: US/10/433,287
CURRENT FILING DATE: 2003-11-12
NUMBER OF SEQ ID NOS: 81
SOFTWARE: PatentIn version 3.2
SEQ ID NO 79
LENGTH: 104644
TYPE: DNA
ORGANISM: homo sapiens
FEATURE:
NAME/KEY: genomic DNA
LOCATION: (1)..(104644)
OTHER INFORMATION: n is an undetermined nucleotide (darp, dcrp, dgrp, or dtrp)
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)..(3)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc feature
LOCATION: (39)..(39)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc feature
LOCATION: (70)..(70)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc feature
LOCATION: (73)..(73)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc feature
LOCATION: (143)..(143)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc feature
LOCATION: (231)..(231)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc feature
LOCATION: (242)..(242)
OTHER INFORMATION: n is a, c, g, or t

```

```

1 FEATURE:
2 NAME/KEY: misc feature
3 LOCATION: (362) .. (362)
4 OTHER INFORMATION: n is a, c, g, or t
5 FEATURE:
6 NAME/KEY: misc feature
7 LOCATION: (11659) .. (11659)
8 OTHER INFORMATION: n is a, c, g, or t
9 FEATURE:
10 NAME/KEY: misc feature
11 LOCATION: (11699) .. (11699)
12 OTHER INFORMATION: n is a, c, g, or t
13 FEATURE:
14 NAME/KEY: MS4A7 initial coding_region
15 LOCATION: (117493) .. (117639)
16 FEATURE:
17 NAME/KEY: MS4A7 coding_region
18 LOCATION: (19439) .. (19573)
19 FEATURE:
20 NAME/KEY: MS4A7 coding_region
21 LOCATION: (21068) .. (21124)
22 FEATURE:
23 NAME/KEY: MS4A7 coding_region
24 LOCATION: (23741) .. (23947)
25 FEATURE:
26 NAME/KEY: MS4A7 coding_region
27 LOCATION: (27037) .. (27158)
28 FEATURE:
29 NAME/KEY: MS4A7 coding_region
30 LOCATION: (28139) .. (28210)
31 FEATURE:
32 NAME/KEY: misc feature
33 LOCATION: (32640) .. (32640)
34 OTHER INFORMATION: n is a, c, g, or t
35 FEATURE:
36 NAME/KEY: MS4A5 initial coding_region
37 LOCATION: (64028) .. (64180)
38 FEATURE:
39 NAME/KEY: MS4A5 coding_region
40 LOCATION: (65145) .. (65277)
41 FEATURE:
42 NAME/KEY: MS4A5 coding_region
43 LOCATION: (66764) .. (66820)
44 FEATURE:
45 NAME/KEY: MS4A5 coding_region
46 LOCATION: (68118) .. (68270)
47 FEATURE:
48 NAME/KEY: MS4A5 coding_region
49 LOCATION: (82002) .. (82109)
50 FEATURE:
51 NAME/KEY: MS4A12 initial coding_region
52 LOCATION: (91488) .. (91760)
53 FEATURE:
54 NAME/KEY: MS4A12 coding_region
55 LOCATION: (95210) .. (95347)
56 FEATURE:
57 NAME/KEY: MS4A12 coding_region
58 LOCATION: (96148) .. (96204)
59 FEATURE:
60 NAME/KEY: MS4A12 coding_region
61 LOCATION: (97665) .. (97981)
62 FEATURE:
63 NAME/KEY: MS4A12 coding_region
64 LOCATION: (100890) .. (101000)
65 FEATURE:
66 NAME/KEY: MS4A12 coding_region
67 LOCATION: (101182) .. (101283)
68 FEATURE:
69 NAME/KEY: misc feature
70 LOCATION: (104561) .. (104625)
71 OTHER INFORMATION: n is a, c, g, or t
72 US-10-433-287-79

```

Query Match 21.9%; Score 30.4; DB 19; Length 104644;  
 Best Local Similarity 53.3%; Pred. No. 89;  
 Matches 64; Conservative 0; Mismatches 56; Indels 0; Gaps 0;  
 Qy 16 GAACTTAATAATGCGCTGCTCACCCCTTTTCTTCAGAAAGAGGCTGACTATTGTCTG 75  
 Db 88420 GAACCCAGATTATTCATGCTAGTCTATAGTTATAGGACTCAGAAATACATTTTACTG 88361  
 Qy 76 GTTATTAACTGTTTATCCCCAAGACCAATATCAAGCTAGACTGTTCTTATTTGTTAA 135  
 Db 88360 GGTCAAGTTCTGTTTATTAAGAAAGATTAAGACAGAGCCAGCCAGTTCTGATTTCTTAA 88301

Search completed: August 4, 2005, 01:55:06  
 Job time : 363 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 3, 2005, 21:47:30 ; Search time 105 Seconds  
(without alignments)  
2166.119 Million cell updates/sec

Title: US-10-070-882A-2

Perfect score: 139  
Sequence: 1 gtgactctgtgcgcagcaact.....ctgtcttctatgttacaaca 139

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents NA:\*  
1: /cgn2\_6/prodata/1/ina/5A.COMB.seq:\*  
2: /cgn2\_6/prodata/1/ina/5B.COMB.seq:\*  
3: /cgn2\_6/prodata/1/ina/6A.COMB.seq:\*  
4: /cgn2\_6/prodata/1/ina/6B.COMB.seq:\*  
5: /cgn2\_6/prodata/1/ina/PCUS.COMB.seq:\*  
6: /cgn2\_6/prodata/1/ina/backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	33	23.7	52494	4	US-09-949-016-16498 Sequence 16498, A
C 2	32.2	23.2	86273	4	US-09-949-016-15273 Sequence 15273, A
C 3	31.8	22.9	102304	4	US-09-949-016-12589 Sequence 12589, A
C 4	30	21.6	601	4	US-09-949-016-81832 Sequence 81832, A
C 5	30	21.6	601	4	US-09-949-016-144837 Sequence 144837, A
C 6	30	21.6	14712	4	US-09-949-016-15840 Sequence 15840, A
C 7	30	21.6	22823	4	US-09-949-016-14121 Sequence 14121, A
C 8	30	21.6	211049	4	US-09-949-016-15770 Sequence 15770, A
C 9	29.8	21.4	301	3	US-09-439-313-293 Sequence 293, App
C 10	29.8	21.4	301	3	US-09-352-616A-293 Sequence 293, App
C 11	29.8	21.4	301	3	US-09-232-149A-293 Sequence 293, App
C 12	29.8	21.4	301	4	US-09-159-812-293 Sequence 293, App
C 13	29.8	21.4	301	4	US-09-636-215-293 Sequence 293, App
C 14	29.8	21.4	301	4	US-09-685-166A-293 Sequence 293, App
C 15	29.8	21.4	301	4	US-09-688-489-293 Sequence 293, App
C 16	29.8	21.4	301	4	US-09-679-426-293 Sequence 293, App
C 17	29.8	21.4	301	4	US-09-759-143-293 Sequence 293, App
C 18	29.8	21.4	301	4	US-09-651-236-293 Sequence 293, App
C 19	29.8	21.4	1140	4	US-09-107-532A-1865 Sequence 1865, App
C 20	29.8	21.4	2143	4	US-09-673-395A-82 Sequence 82, App
C 21	29.8	21.4	3174	4	US-09-489-847-86 Sequence 86, App
C 22	29.2	21.0	16593	3	US-08-961-527-52 Sequence 52, App
C 23	29	20.9	3010	3	US-08-714-918-71 Sequence 71, App
C 24	29	20.9	3010	3	US-09-265-315-71 Sequence 71, App
C 25	29	20.9	3010	3	US-09-265-315-71 Sequence 71, App
C 26	29	20.9	3010	3	US-09-266-417-71 Sequence 71, App
C 27	29	20.9	3010	4	US-09-528-709-71 Sequence 71, App

28	29	20.9	3010	4	US-09-527-745-71 Sequence 71, App
29	29	20.9	194790	4	US-09-949-016-15393 Sequence 15393, A
30	29	20.9	194889	4	US-09-949-016-15654 Sequence 15654, A
C 31	28.8	20.7	601	4	US-09-949-016-47920 Sequence 47920, A
C 32	28.8	20.7	2347	1	US-08-453-695A-113 Sequence 113, App
C 33	28.8	20.7	2347	1	US-08-268-161A-113 Sequence 113, App
C 34	28.8	20.7	2347	2	US-08-453-702A-113 Sequence 113, App
C 35	28.8	20.7	2347	5	US-09-099-639-113 Sequence 113, App
C 36	28.8	20.7	2347	5	PCT-US95-08071-113 Sequence 113, App
C 37	28.6	20.6	601	4	US-09-949-016-40625 Sequence 40625, A
C 38	28.6	20.6	601	4	US-09-949-016-40626 Sequence 40626, A
C 39	28.6	20.6	601	4	US-09-949-016-148424 Sequence 148424, A
C 40	28.6	20.6	601	4	US-09-949-016-148425 Sequence 148425, A
C 41	28.6	20.6	2676	3	US-09-212-971-11 Sequence 11, App
C 42	28.6	20.6	2676	3	US-08-800-929A-11 Sequence 11, App
C 43	28.6	20.6	2676	3	US-09-617-053A-11 Sequence 11, App
C 44	28.6	20.6	16397	4	US-08-956-171E-205 Sequence 205, App
C 45	28.6	20.6	16397	4	US-08-781-986A-205 Sequence 205, App

## ALIGNMENTS

RESULT 1  
US-09-949-016-16498/c  
Sequence 16498, Application US/09949016  
Patent No. 6812339  
GENERAL INFORMATION:  
APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
FILE REFERENCE: C1001307  
CURRENT APPLICATION NUMBER: US/09/949,016  
CURRENT FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 16498  
LENGTH: 52494  
TYPE: DNA  
ORGANISM: Human  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1) - (52494)  
OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-16498  
Query Match 23.7%; Score 33; DB 4; Length 52494;  
Best local similarity 65.8%; Pred. No. 0.66;  
Matches 48; Conservative 0; Mismatches 25; Indels 0; Gaps 0;  
QY 18 ACTAATAATAGCTGCTCACCCTTTCTTCAGAAAGGTCACATTTGCTGCT 77  
DB 11879 ATTGAATATCTGCTGCTTTTCTTCTTTTACCAATATGCTATTAATGATCATTT 11820  
QY 78 TTTATTAATGCTTT 90  
DB 11819 TTTATTAATGCTTT 11807  
RESULT 2  
US-09-949-016-15273/c  
Sequence 15273, Application US/09949016  
Patent No. 6812339  
GENERAL INFORMATION:  
APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF

```

1 FILE REFERENCE: CL001307
2
3 CURRENT APPLICATION NUMBER: US/09/949,016
4
5 CURRENT FILING DATE: 2000-04-14
6
7 PRIOR APPLICATION NUMBER: 60/241,755
8
9 PRIOR FILING DATE: 2000-10-20
10
11 PRIOR APPLICATION NUMBER: 60/237,768
12
13 PRIOR FILING DATE: 2000-10-03
14
15 PRIOR APPLICATION NUMBER: 60/231,498
16
17 PRIOR FILING DATE: 2000-09-08
18
19 NUMBER OF SEQ ID NOS: 207012
20
21 SOFTWARE: PatSeq for Windows Version 4.0
22
23 SEQ ID NO 15273
24
25 LENGTH: 86273
26
27 TYPE: DNA
28
29 ORGANISM: Human
30
31 US-09-949-016-15273

```

Query Match	23.2%	Score 32.2	DB 4	Length 86273
Best Local Similarity	54.7%	Pred. No. 1.6		
Matches	64	Conservative 0	Mismatches 53	Indels 0
			Gaps 0	
Qy	23	AATAATGCGCGCCACCCCTCTTTCTTCAGAAAAGAGGAGACATTTGCTGTATT	82	
Db	45531	AATAACTGTGTACCAAAACTATTTTCACTGGAGAGATGAGACAATTTTCTGTCACT	45477	
Qy	83	AACGTTTATCCCAAAAGACACATAATCAACGCTAGACTGTTCTATTGTTAACACA	139	
Db	45471	TACATTTTATTTTCATATCATATTTTCCAAAGAGAGAGTTGGGATTTTAAAAAATA	45415	

```

RESULT 3
US-09-949-016-12589/C
; Sequence 12589, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12589
; LENGTH: 102304
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(102304)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12589

```

	Query Match	Similarity	%	Score	DB	Length
Best Local	54	Conservative	0	Mismatches	37	Gaps
Matches	54	Conservative	0	Mismatches	37	Gaps
Query	20	TTAATATATGCTGCTACCCCTCTTTCTTCAGAAAGGATCACTTTGTCTGTT	79			
DB	6060	TTTACCACTCCAAATCATCTCTCTTGCCAGAAATTATTAATATTTGTCACTTG	6001			
Query	80	ATTAACGTGTTATCCCAAGCACAATATC	110			
DB	6000	ATTAGTACTAATTGACAGAGCAGAGATAC	5970			

## RESULT 4

```

US-09-949-016-81832/c
; Sequence 81832 Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,766
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 81832
;
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-81832

```

Query Match	21.6%	Score 30	DB 4	Length 601
Best Local Similarity	54.5%	Pred. No. 1.1		
Matches	60	Conservative	0	Mismatches 50
			Indels	0
			Gaps	0

  

QY	1	GTGACTCTGTGTCGACAGCAACTTAAATTAATCCCTGTGCTACACCTCTTTTCTTGAGAAAGG	60
DB	154	GAGCACTTGGGCAAAATCACTTAATCTGTCTGTGCTCAGTTTCTTATCTATTAATAATGGG	95
QY	61	GTGACTATTTGTCTGGTTATTAACTGTTATATCCCAAGCACCATAATC	110
DB	94	GAGGATATATGGTGTCAACTTTTCTTTATATAAGATTAAATTAAGATTATGC	45

```

RESULT 5
US-09-949-016-144837/c
; Sequence 144837, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949, 016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 144837
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-144837

```

	Query Match	Similarity	21.6%	Score 30;	DB 4;	Length 601;
	Best Local	Similarity	54.5%	Pred. NO. 1.1;		
	Matches	60;	Conservative	0;	Mismatches	50; Indels 0; Gaps 0;
Qy	1	GTGACTCTGTGGCGAGCACTTAATAATAGCCCTGCGCTACCCCTCTTTTCTTGAGAAAGG	60			
Db	154	GAGACTTGTGGGCGAATACACTTAATCTGTCTGTGCGCTCAGTTCTCTTAATCTATAAATGGG	95			
Qy	61	GTGACTATTGTCTGTATTATTAAGTGTTTATATCCCAAGACCACTAATC	110			
Db	94	GAGATAATATGGTGTCAACTCTTTTACTTTATATAAGATTAATTAAGTTATGTC	45			

## RESULT 4

RESULT 6  
US-09-949-016-15840  
; Sequence 15840, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:

; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 15840  
; LENGTH: 14712  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-15840

Query Match 21.6%; Score 30; DB 4; Length 14712;  
Best Local Similarity 54.5%; Pred. No. 4.4;  
Matches 60; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 1 GTGACTTGTGTCGACCACTTAATATATCCCTGCTCACCCTTTCTTTCAGAAAGAG 60  
DB 3243 GAGACTTTGGCGAAATCACTTATCTGTGCTGCTCAGCTTTCCTATCTATAAATGAG 3302  
QY 61 GTGACTTGTGTCGCTGTTTATTAAGTTCATCCCAAGCAGCATATC 110  
DB 3303 GAGAAATATGCTGTCACCTTTTACTTTATTAAGATTAAATAGATGATC 3352

RESULT 7  
US-09-949-016-14121  
; Sequence 14121, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:

; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 14121  
; LENGTH: 22823  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-14121

Query Match 21.6%; Score 30; DB 4; Length 22823;  
Best Local Similarity 54.5%; Pred. No. 5.3;  
Matches 60; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 1 GTGACTTGTGTCGACCACTTAATATATGCTGCTCACCCTTTCTTTCAGAAAGAG 60  
DB 3243 GAGACTTTGGCGAAATCACTTATCTGTGCTGCTCAGCTTTCCTATCTATAAATGAG 3302

QY 61 GTGACTTGTGTCGCTGTTTATTAAGTTCATCCCAAGCAGCATATC 110  
DB 3303 GAGAAATATGCTGTCACCTTTTACTTTATTAAGATTAAATAGATGATC 3352

RESULT 8  
US-09-949-016-15770/c  
; Sequence 15770, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:

; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 15770  
; LENGTH: 211049  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-15770

Query Match 21.6%; Score 30; DB 4; Length 211049;  
Best Local Similarity 61.5%; Pred. No. 14;  
Matches 48; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

QY 13 GAGCACTTAATTAATATGCTGCTCACCCTTTCTTTCAGAAAGAGGCTGACTATTTGT 72  
DB 164239 GAGGCCCTTACTTCTTATCTCATCATCCTTATCTCAATATGAAAGATTCATTTTC 164240  
QY 73 CTGCTTATTAAGTCTTT 90  
DB 164239 ATTGTTGTTTAATTAATTT 164222

RESULT 9  
US-09-439-313-293  
; Sequence 293, Application US/09439313  
; Patent No. 6329505  
; GENERAL INFORMATION:

; APPLICANT: Xu, Jiangchun  
; APPLICANT: Dillon, Davin C.  
; APPLICANT: Mitcham, Jennifer L.  
; APPLICANT: Harlocker, Susan Louise  
; APPLICANT: Jiang Yuqun  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Kalos, Michael  
; APPLICANT: Fanger, Gary  
; APPLICANT: Retter, Mark  
; APPLICANT: Solk, John  
; APPLICANT: Day, Craig  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND  
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER  
; FILE REFERENCE: 210121.427C9  
; CURRENT APPLICATION NUMBER: US/09/439,313  
; CURRENT FILING DATE: 1999-11-12  
; NUMBER OF SEQ ID NOS: 575  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 293  
; LENGTH: 301  
; TYPE: DNA  
; ORGANISM: Homo sapien  
US-09-439-313-293

Query Match 21.4%; Score 29.8; DB 3; Length 301;



```
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-636-215-293

Query Match
Best Local Similarity 21.4%; Score 29.8; DB 4; Length 301;
Matches 55; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY 43 CTTTCTTCAGAAAGGGTGACTATTGTCTGTTTATTAACTGTTATCCCAAGCA 102
DB 30 CTGTTCTCACTGAAAGTCTGGCTAATGCTCTGTGTAGTCACTTCTGATTCGACAAATC 89

QY 103 CCATAATCAAGCGCTAGACTGTTCTTATTGTTAAACA 139
DB 90 AATCAATCAATGGCCTTAGAGCACTGACTGTTAAACA 126

RESULT 14
US-09-685-166A-293
; Sequence 293, Application US/09685166A
; Patent No. 6630305
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqul
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.427C21
; CURRENT APPLICATION NUMBER: US/09/685,166A
; NUMBER OF SEQ ID NOS: 898
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 293
; LENGTH: 301
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-685-166A-293

Query Match
Best Local Similarity 21.4%; Score 29.8; DB 4; Length 301;
Matches 55; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY 43 CTTTCTTCAGAAAGGGTGACTATTGTCTGTTTATTAACTGTTATCCCAAGCA 102
DB 30 CTGTTCTCACTGAAAGTCTGGCTAATGCTCTGTGTAGTCACTTCTGATTCGACAAATC 89

QY 103 CCATAATCAAGCGCTAGACTGTTCTTATTGTTAAACA 139
DB 90 AATCAATCAATGGCCTTAGAGCACTGACTGTTAAACA 126

RESULT 15
US-09-688-489-293
; Sequence 293, Application US/09688489
; Patent No. 6664377
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer Lynn
```

```
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE
; FILE REFERENCE: 210121.427D2
; CURRENT APPLICATION NUMBER: US/09/688,489
; NUMBER OF SEQ ID NOS: 338
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 293
; LENGTH: 301
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-688-489-293

Query Match
Best Local Similarity 21.4%; Score 29.8; DB 4; Length 301;
Matches 55; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY 43 CTTTCTTCAGAAAGGGTGACTATTGTCTGTTTATTAACTGTTATCCCAAGCA 102
DB 30 CTGTTCTCACTGAAAGTCTGGCTAATGCTCTGTGTAGTCACTTCTGATTCGACAAATC 89

QY 103 CCATAATCAAGCGCTAGACTGTTCTTATTGTTAAACA 139
DB 90 AATCAATCAATGGCCTTAGAGCACTGACTGTTAAACA 126

Search completed: August 4, 2005, 01:48:59
Job time : 107 secs
```

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 3, 2005, 21:27:30 ; Search time 6694 Seconds  
(without alignments)  
1006.166 Million cell updates/sec

Title: US-10-070-882A-2

Perfect score: 139

Sequence: 1 gtgactctgtgcgcacgaact.....ctgtctctatgttaacaca 139

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl.\*  
1: gb\_da.\*  
2: gb\_hg.\*  
3: gb\_in.\*  
4: gb\_cm.\*  
5: gb\_ov.\*  
6: gb\_pac.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_str.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	139	100.0	139	6	AX100259 Sequence
2	139	100.0	990	1	STYPHOPA M25241 S.typhimuri
3	139	100.0	2459	1	AL272210 Salmonell
4	134.2	96.5	254050	1	AL627269 Salmonell
5	134.2	96.5	300029	1	AE016839 Salmonell
6	129	92.8	2150	1	M24424 S.typhimuri
7	113	81.3	24578	1	STYPHOPA AE008754
8	86	61.9	22182	1	AE008753 Salmonell
9	40	28.8	41	6	AX17848 Sequence
10	36.6	26.3	166554	10	AC110381
11	36.4	26.2	11716	1	AE010854 Methanosa
12	36	25.9	44532	9	AC109830 Homo sapi
13	35.6	25.6	245108	2	AC103069 Rattus no
14	35.4	25.5	162561	2	AC113887 Rattus no
15	35	25.2	126477	8	AC146720 Medicago
16	34.8	25.0	140884	2	AL357116 Homo sapi
17	34.8	25.0	182553	2	AC025386 Homo sapi
18	34.8	25.0	182553	9	AL449266 Human DNA
19	34.8	25.0	187214	2	AL354724 Homo sapi

20	34.8	25.0	196175	10	AC068501 Mus muscu
21	34.8	25.0	212789	10	AC124676 Mus muscu
22	34.6	24.9	74999	9	AC008960 Homo sapi
23	34.6	24.9	155025	9	AC034246 Homo sapi
24	34.6	24.9	171274	2	AC026476 Homo sapi
25	34.6	24.9	183376	2	AC129621 Rattus no
26	34.6	24.9	184118	9	AC016580 Homo sapi
27	34.6	24.9	192676	10	AL671706 Mouse DNA
28	34.6	24.9	216882	2	AC132716 Rattus no
29	34.6	24.9	234105	10	AC091606 Mus muscu
30	34.4	24.7	192986	2	CR352234 Dantio rer
31	34.2	24.6	117189	2	AC147547 Rattus no
32	34.2	24.6	220737	2	AC128459 Rattus no
33	34.2	24.6	222801	2	AC093951 Rattus no
34	33.6	24.2	82279	9	AL391071 Human DNA
35	33.6	24.2	192518	2	HSBA539A6
36	33.6	24.2	192992	10	AL691417 Mus muscu
37	33.4	24.0	240887	10	AC102562 Mus muscu
38	33.4	24.0	242893	10	AC102550 Mus muscu
39	33.2	23.9	552	9	HSB324614
40	33	22.7	35869	9	AC144561 Homo sapi
41	33	23.7	165824	10	AL663090 Mouse DNA
42	33	23.7	343837	2	AC102887 Mus muscu
43	32.8	23.6	152435	10	AC121566 Mus muscu
44	32.8	23.6	157392	5	BX548028 Zebrafish
45	32.6	23.5	65377	9	AL391356 Human DNA

#### ALIGNMENTS

RESULT 1	AX100259	Sequence 2 from Patent WO0119974.	139 bp	DNA	linear	PAT 02-APR-2001
LOCUS	AX100259	Sequence 2 from Patent WO0119974.	139 bp	DNA	linear	PAT 02-APR-2001
DEFINITION	AX100259	Sequence 2 from Patent WO0119974.	139 bp	DNA	linear	PAT 02-APR-2001
ACCESSION	AX100259	Sequence 2 from Patent WO0119974.	139 bp	DNA	linear	PAT 02-APR-2001
VERSION	AX100259.1	GI:13539143				
KEYWORDS						
SOURCE						
ORGANISM						
Salmonella typhimurium						
Salmonella typhimurium						
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Salmonella.						
REFERENCE						
1	Titball, R.W. and Bullifant, H.L.					
AUTHORS						
TITLE						
JOURNAL						
The Secretary of State for Defence (GB)						
Location/Qualifiers						
1.139						
/organism="Salmonella typhimurium"						
/mol_type="unassigned DNA"						
/db_xref="taxon:602"						
ORIGIN						
Query Match	100.0%	Score 139;	DB 6;	Length 139;		
Best Local Similarity	100.0%;	Pred. No. 7.3e-29;				
Matches 139;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;		
1	GTGACTCTGTGCGACGACCACTTAATATATGCTGCTCACCCTCTTTCTTCAGAAAGAGG	60				
1	GTGACTCTGTGCGACGACCACTTAATATATGCTGCTCACCCTCTTTCTTCAGAAAGAGG	60				
1	GTGACTCTGTGCGACGACCACTTAATATATGCTGCTCACCCTCTTTCTTCAGAAAGAGG	60				
61	GTGACTCTGTGCGACGACCACTTAATATATGCTGCTCACCCTCTTTCTTCAGAAAGAGG	120				
61	GTGACTCTGTGCGACGACCACTTAATATATGCTGCTCACCCTCTTTCTTCAGAAAGAGG	120				
61	GTGACTCTGTGCGACGACCACTTAATATATGCTGCTCACCCTCTTTCTTCAGAAAGAGG	120				
121	TGTTCTTATTTGTTAAACA 139					
121	TGTTCTTATTTGTTAAACA 139					
121	TGTTCTTATTTGTTAAACA 139					
RESULT 2						
STYPHOPA						

LOCUS	STYHOPA	990 bp	DNA	linear	BCT 26-APR-1993							
DEFINITION	S.typhimurium phoP gene encoding PhoP virulence protein, complete cds.											
ACCESSION	M25241	GI:154263										
VERSION	M25241.1											
KEYWORDS	phoP gene; virulence protein.											
SOURCE	Salmonella typhimurium											
ORGANISM	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Salmonella.											
REFERENCE	1 (bases 1 to 990)											
AUTHORS	Grisman, E.A., Chiao, E., Lippe, C.J. and Heffron, F.											
TITLE	Salmonella typhimurium phoP virulence gene is a transcriptional regulator											
JOURNAL	Proc. Natl. Acad. Sci. U.S.A.	86 (18)	7077-7081	(1989)								
MEDLINE	89386683											
PUBMED	2674945											
COMMENT	Original source text: S.typhimurium (strain 140286) DNA, clone PB5381. Draft entry and computer-readable sequence [1] kindly submitted by F. Heffron, 01-JUN-1989.											
FEATURES	Location/Qualifiers											
source	1..990											
	/organism="Salmonella typhimurium"											
	/mol_type="genomic DNA"											
	/db_xref="taxon:602"											
RBS	140..142											
	/note="ribosome binding site"											
CDS	151..825											
	/note="phoP protein"											
	/codon_start=1											
	/transl_table=11											
	/protein_id="AAA27187.1"											
	/db_xref="GI:154264"											
	/translation="MMRLVIVVEDNALLRHLKVLQDSGHQVDAADAEADYYLNHSLPDAIVDIGLPDEGSLIRWRSSDVSLEVLTLAREGMQDVVEVLSGADQVYRKPFHFEVNAQMQLMRNRSGASQVINPPQVDLSRELSVNEVILKTRFETIHEYLIRNGKVSFKDSIMLQLYPDABLRSHSTIDVLMGRKKIKQAOYPHDVTIVRGQGLFLEIR"											
ORIGIN	25 minutes on S.typhimurium map.											
Query Match	100.0%; Score 139; DB 1; Length 990;											
Best Local Similarity	100.0%; Pred. No. 5.2e-29;											
Matches 139; Conservative	0; Mismatches 0; Indels 0; Gaps 0;											
Oy	1 GTGACTCTGTGTCGACGAAGCTTAATAAAGCCGCTCAACCTCTTTCTTCAGAAAGAG 60											
Db	1 GTGACTCTGTGTCGACGAAGCTTAATAAAGCCGCTCAACCTCTTTCTTCAGAAAGAG 60											
Oy	61 GTGACTATTTGTCTGTTTATTACTGTTTATCCCAAGACACATATCAACGCTAGAC 120											
Db	61 GTGACTATTTGTCTGTTTATTACTGTTTATCCCAAGACACATATCAACGCTAGAC 120											
Oy	121 TGTCTCTATTTGTTAACACA 139											
Db	121 TGTCTCTATTTGTTAACACA 139											
RESULT 3												
LOCUS	STY272210	2459 bp	DNA	linear	BCT 30-NOV-2001							
DEFINITION	Salmonella typhimurium phoP gene and phoQ gene, strain SJ1344.											
ACCESSION	AJ272210											
VERSION	AJ272210.1	GI:7007368										
KEYWORDS	membrane sensor protein, phoP gene; phoQ gene; transcriptional regulator; virulence gene.											
SOURCE	Salmonella typhimurium											
ORGANISM	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Salmonella.											
REFERENCE	1											
AUTHORS	Cano, D.A., Martinez-Moya, M., Pucciarelli, M.G., Groisman, E.A., Casadesus, J. and Garcia-Del Portillo, F.											

FEATURES	source
JOURNAL	Salmonella enterica serovar Typhimurium response involved in
LINE	attenuation of pathogen intracellular proliferation
REFERENCE	Infect. Immun. 69 (10), 6463-6474 (2001)
AUTHORS	2 (bases 1 to 2459)
TITLE	Garcia-del Portillo,F.
JOURNAL	Direct Submission
	Submitted (16-FEB-2000) Garcia-del Portillo F., De Biología
	C.S.I.C.-Universidad Autónoma de Madrid, Centro de Biología
	Molecular 'Severo Ochoa', Campus de Cantoblanco, 28049 Madrid,
	SPAIN
	Location/Qualifiers
	1. 2459
	/organism="Salmonella typhimurium"
	/mol_type="genomic DNA"
	/strain="SL1344"
	/db_xref="taxon:602"
	240 . 914
	/gene="phop"
	240 . 914
	/gene="phop"
	/function="virulence transcriptional regulator"
	/note="member of a two-component regulatory system"
	/codon_start=1
	/transl_table=11
	/product="Phop protein"
	/protein_id="CAB75591.1"
	/db_xref="GI:7007369"
	/db_xref="GOA:P14146"
	/db_xref="UniProt/Swiss-Prot:P14146"
	/translation="MMRVLYVEDNALIRHLKVOLDGCHVDAAEDARADYVINEH
	LPDIAVDGLPDGDLGLRRSSDVSIPVLTLRBGMQKVEVTLGGADYVTK
	LFPIEYVMQALMRNSGLAQVINIPFQVLSRELISVNEVTKLTFEYITWE
	TLIRNGKVVSKDLSMLQIPLPDARELRSHSDIVLGMGLRKIKIQAYPHDIVTVRGGQ
	YLFELR"
	914 . 2377
	/gene="phoQ"
	914 . 2377
	/gene="phoQ"
	/function="membrane sensor protein"
	/note="member of a two-component regulatory system"
	/codon_start=1
	/transl_table=11
	/product="PhoQ protein"
	/protein_id="CAB75592.1"
	/db_xref="GI:7007370"
	/db_xref="GOA:P14147"
	/db_xref="UniProt/Swiss-Prot:P14147"
	/translation="MNFARHFLPLSRPFLATNAGVVLVSLATGVALVGSVSF
	DKTFLRGESNLFYTLARWENKKISVLPBNDIMQSPMTLLIYDGTGLKLTQNTN
	PWLKISPEMLKINGFHEIETNVDAITLSBDSHQBKLKEVRDDDLAEIHWSSVA
	VNIIPASAPMPLTIVVDITPIELKRSYMWQFVYVLAANLLVPLMIAMWSI
	RPFLAIRRELELDHHRMLNPETRELISLVANLQLLKSEERXKRYKTTLDLIL
	HSKLPIAVLQSTIRSLRNKMSYKSKAPMLBQISITISQOIGYVLRASRGSGVLL
	SREHLPAPLNDNLISALNRYQKQGVNISNDISPEISFGEQNDPVEWGNVINDAC
	KYCLIEFVEISARQTDHHLFVEDDGGIAPHSKSLVFDGGQADTLTPQGVGLAVVA
	REIRFQYAGQIIASDSLGGARMEVVGROHPTQKE"



misc\_feature

/db\_xref="UniProt/TREMBL:082787"  
/translation="MATTSLRFLTHPEVKSRRGIGLTHALADISGLAFDRIFWTSDDGT  
PRTAQFPPMVRFPSPPLDGLHLTAIPDSSALVTRTDTTPDAAPRWGNHFTAVA  
FAITNOISGFPSRDVQLCVGPOLTRAYKRNNAVLPAGDYPYLITBAALDLIOQ  
RPDAGIOMGSPFNVLGVSAAMEEDSMYLVLRGDVIFPVNVCSCSICITTVSPKGO  
KHSGEPLATLICAPRTAODNGVDVPGDNQIAANSNPVIRGDEVILATAPRAYGTT  
LDNSTPRHPGSSTIVMGQCOTPCNNQNVLLBQLENGGIRIPYSRCAGICGCCRI  
RLSEGVSPILKSNAGMDGGTTLGSCVPKLTALDEN"  
complement(4737..4955)  
/gene="STY1081"  
/note="Pfam match to entry PF00111 ferZ, 2Fe-2S  
iron-sulfur cluster binding domains, score 55.90, E-value  
8.5e-13"

Gene

CDS

5924..8032  
/gene="STY1082"  
5924..8032  
/gene="STY1082"  
/note="Similar to Haemophilus influenzae hypothetical  
protein h10116 H10116 SW:YCBY\_HAEIN (P44524; P43945) faeta  
access: E() = 0, 60.3% id in 710 aa  
Orthologue of E. coli YCBY\_ECOLI; Faeta hit to YCBY\_ECOLI  
(702 aa), 92% identity in 702 aa overlap"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="CAD08187.1"  
/db\_xref="GI:16502236"  
/db\_xref="GOA:Q82786"  
/db\_xref="UniProt/TREMBL:082786"  
/translation="MNSLFASARGALEELIKETLEGAVGCGVGHPOGDTRLI  
YOSIMMSRLASRIILPMGRCKVSDLIIDFGVAIIMWTIPNGARFPAHDSSGNTTI  
RNSOYAMAKYKDAIYDAFTRKULPNRVDBESDFDIINWINKETSIALDLSGDSLH  
LNGVRTGLAPIKETLAAILVMRSQMOPETPLDMDCSGTILLISAAMATIDRAGL  
HGHWMPGSOADEAIWOEVAAEQTRARKGLAESHSFHYGSDSAPARIERARSNA  
RAGIGSILIFEVEDVALONSPLEKPYGVIVISNPGEISDPDALIALSHLIGRTMK  
NOFGNNLILFSASPDLGSLQLRADRGKAKNGPDCVOKNHIAETALSDEKPATVA  
EOYANLRKNIKKEKRAOEIGIECRVADADIPEYNNAVDRGMAVIVOEAPPTV  
DKAKAKORLPDIIATLVSIGIPPKVLVKTBRQKQKQKQYQMSKSEFEVSRINA  
RLMVNLITLDLDELFDNRILARRMLDEMSGKQPLNFSTGSASVHAEGGAKRNTT  
VMSRYLTMABRENLDLNGISGRAHRLIQADCGLMRLEANEPDILFIIDPFSSNK  
MESSEPVQCDHVHALMDLRLRLKKGITIMEFNSNNKRGRFMDDLGLAELGHTAOETIXT  
LSPDFARNROIHNCWILRAA"

misc\_feature

5927..7060  
/gene="STY1082"  
/note="Pfam match to entry PF01170 UPF0020,  
Uncharacterized protein family UPF0020, score 781.30,  
E-value 3.7e-231"  
6506..6541  
/gene="STY1082"  
/note="PB01261 Uncharacterized protein family UPF0020  
signature  
8045..9952  
/gene="STY1083"  
8045..9952  
/gene="STY1083"  
/note="Faeta hit to YBIT\_ECOLI (530 aa), 32% identity in  
526 aa overlap  
Faeta hit to YJUK\_ECOLI (554 aa), 34% identity in 524 aa  
overlap

Gene

CDS

8045..9952  
/gene="STY1083"  
8045..9952  
/gene="STY1083"  
/note="Faeta hit to YBIT\_ECOLI (530 aa), 32% identity in  
526 aa overlap  
Faeta hit to YJUK\_ECOLI (554 aa), 34% identity in 524 aa  
overlap

Key Match

Seq Local Similarity 97.8%; Pred. No. 4.7e-26;  
Matches 136; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

1 GTGACTCTGGTCACAGCACTTAATAATATGCTGCCTCACCCCTCTTTCTTCAGAAAGAGG 60  
175998 GTGAATCTGGTCACAGCACTTAATAATATGCTGCCTCACCCCTCTTTCTTCAGAAAGAGG 175999  
61 GTGACTATTTGCTGCTTTATTAACGTTTATCCCCAAGACCATATATCAAGCGTAGAC 120  
175938 GTGACTATTTGCTGCTTTATTAACGTTTATCCCCAAGACCATATATCAAGCGTAGAC 175939  
175938 GTGACTATTTGCTGCTTTATTAACGTTTATCCCCAAGACCATATATCAAGCGTAGAC 175939

Db	Accession	Source	Gene	CDS
07	121	TGTTCTTAATGTTACACA	139	
07	175878	TGTTCTTAATGTTACACA	175860	
RESULT 5				
LOCUS	AB016839	300029 bp DNA	linear	BCT 21-MAR-2003
DEFINITION	Salmonella enterica subsp. enterica serovar Typh1 Ty2, section 6 of 16 of the complete genome.			
ACCESSION	AB016839			
VERSION	AB016839.1			
KEYWORDS	GI:29137536			
SOURCE	Salmonella enterica subsp. enterica serovar Typh1 Ty2			
ORGANISM	Salmonella enterica subsp. enterica serovar Typh1 Ty2			
REFERENCE	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Salmonella.			
AUTHORS	1 (bases 1 to 300029) Deng, W., Lioy, S.-R., Plunkett III, G., Mayhew, G.F., Rose, D.J., Burland, V., Kodoyianni, V., Schwartz, D.C. and Blattner, F.R.			
TITLE	Comparative Genomics of Salmonella enterica Serovar Typh1 Strains Ty2 and CT18			
JOURNAL	J. Bacteriol. 185 (7), 2330-2337 (2003)			
MEDLINE	22531367			
PUBMED	12644504			
REFERENCE	2 (bases 1 to 300029) Deng, W., Lioy, S.-R., Plunkett, G. III, Mayhew, G.F., Rose, D.J., Burland, V., Kodoyianni, V., Schwartz, D.C. and Blattner, F.R.			
AUTHORS	Direct Submission			
TITLE	Submitted (25-SEP-2002) Laboratory of Genetics, University of Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA			
JOURNAL	Location/Qualifiers			
FEATURES	1..300029			
source	/organism="Salmonella enterica subsp. enterica serovar Typh1 Ty2"			
	/mol_type="genomic DNA"			
	/strain="Ty2"			
	/serovar="Typh1"			
	/sub_species="enterica"			
	/db_xref="taxon:209261"			
	245..1336			
	/locus_tag="tl459"			
	245..1336			
	/locus_tag="tl459"			
	/locus_tag="tl459"			
	/note="corresponds to STY1522 from Accession AL513382: Salmonella typh1 CT18"			
	/codon_start=1			
	/transl_table=1			
	/product="putative secreted hydrolase"			
	/protein_id="AA069099.1"			
	/db_xref="GI:29137537"			
	/translation="MMKSKRMLEFIAPALFSYAMAACCTTALIQKQDI FQRTGLE YMDLPMLTYYPAGTQFDKTFDGSQGISYQAKYPLATSTITDDSDILLEGMS AGLSFNSMIMNAQLPLPASEYKQALPTVLSLGMALAPATVGEVQAKKEGFMSR ELHFRGDLRSCTGCIKXNQPDGIAIALDPSGTSVSRIPVNGVATPAFLNLNNK LYTQJLNNDRSGCTGCIKXNQPDGIAIALDPSGTSVSRIPVNGVATPAFLNLNNK DANMTLHINSRPRPKVITTYDYGSGSGEGNATKPKVSETVWTTLSDLTHGEMVNG YNDINVTWLSLQPKNAIAPFEKINVG"			
	1382..2173			
	/locus_tag="tl460"			
	1382..2173			
	/locus_tag="tl460"			
	/note="corresponds to STY1521 from Accession AL513382: Salmonella typh1 CT18"			
	/codon_start=1			
	/transl_table=1			
	/product="putative regulatory protein"			
	/protein_id="AA069100.1"			
	/db_xref="GI:29137538"			
	/translation="METHGGCFSLCTVYNHDKQDDMLNPKPIWVIGRMVLR /EVRERLYQKTLANLILKINDNIFPPGFLPPEBELAKQUGVSASRLREALVLEIGS WIVOSGNGVIVSDKRLHSDYITTEILSTRELVDSCARLAQNDVANDVINOIEAVY HIRMOAINDNNVHGPFYSLDKFPHLAISEASRNLVLPMSRMLMORINIPVAGIDEGS			

gene  
GDRVLTNLNKKHAIYDAMRQSDASAYEGSLHRLRYRKIVGG"  
complement (2232. .3257)  
/locus\_tag="t1461"  
complement (2232. .3257)  
/locus\_tag="t1461"  
/note="corresponds to STY1520 from Accession AL513382:  
Salmonella typhi CT18"  
/codon\_start=1  
/transl\_table=11  
/product="putative alcohol dehydrogenase"  
/protein\_id="AAO69101.1"  
/db\_xref="GI:29137539"  
/translation="MKKAVFINKRGVEVTRWVDYHPHKKEHYLIKYDAAGICGSDIG  
AFRCNIPVYTPRIILGHVYGVIVDSEGTMPDNIKKGDRIYDPIYVYCGHCYCSGR  
TNCCENLIVCVIHIDGAMQEVVTHPAHLIKKIPNNVSEKAPLAEPLTTLAALHPAN  
VKAGGVYAIIDGAGIAGLMAALSARHAYATPLIDIVBERLRVQKLGVPVYNLAADQ  
VIDAKNTHGTGMAOVIVTASGANSARNTLDIASFEGRISFGMGKQERTSLPTNLIIT  
KVELDRSRRISAGEPDEALMLSTLEINPDVYSKVNLDEIDPAVKELDRPERYL  
KINAVFH"  
complement (3232. .4518)  
/locus\_tag="t1462"  
complement (3232. .4518)  
/locus\_tag="t1462"  
/note="corresponds to STY1519 from Accession AL513382:  
Salmonella typhi CT18"  
/codon\_start=1  
/transl\_table=11  
/product="membrane transport protein"  
/protein\_id="AAO69102.1"  
/db\_xref="GI:29137540"  
/translation="MKRKTGMVLVPLFLFYTHLVNDRSALSTAPLIEKEGFNAA  
EKGMIKFSAPFGIYALFNFDGMSADKRGPKVFILIALMSVFGGLGTGTGLTMIL  
VRLVLEGMEGEPVSAAGNKIINNNISRKESALAIIGESAGPILGAVSGPVLGALLATP  
GRRPAPGIIIFGLVWVNLVWFYIVSDKPTMSKRLAPEERIDFENHEDVLIISDGRAPY  
SIGWYKGVQSMVATTLAFESYNIILFPLTFWPSYLNHSLHLDIKESIVATVYPIWG  
AIGMLVGDGQSVYVIRIGGNALSRLLIGVCTAGAAVCAVSGVSTIGSALTMSV  
SIFLLYLTGPITYMAVITQDVHKKDYGSYGAAMHGLANIISGILGYLTGFIYVQSGKID  
YAFYLAGAIAIVSSILVFPVFKSGFKANESQSCVH"  
complement (4695. .6251)  
/gene="pqaa"  
/locus\_tag="t1463"  
complement (4695. .6251)  
/gene="pqaa"  
/locus\_tag="t1463"  
/note="corresponds to STY1518 from Accession AL513382:  
Salmonella typhi CT18"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAO69103.1"  
/db\_xref="GI:29137541"  
/translation="MKRKYLMKLCMLAFALVKGESMKIYLVIVYLFISTKYVTL  
HNINIFCRNSPCDLSHLPDVRBOISGTPAKTLLISTAPLAOVYAHVHLISQMSRP  
DMVTPAQRRNHVDIYIPETAKEHIALVANNNGINYEKGIOIPSKVPDPPQOTASJA  
RPLTNIVIVSVDIPNOYLTIPQDKKPLKEDSVSRSMALFMEAEQELMPLNIPWT  
ALSQMARLAKELTQOMINSPFIITGISRGGTWTMLADPVEAIVPAILDILDDIDA  
SLIEHYISGQWMPITFYPYQOQIDEKIKSPTFOLAOIIDPRVANTVYQPLAP  
KTIIVASGDGFPVNTREPVYSKIPGVKSLIVNNMHSINOFABESLVDFIRFOS  
KTLIPOLIELIHMLITVFESEAPYKAVYRWANPNARDPRYAGIGIYQPLTIDIPAN  
NKISTLNSPKGMEATYIETATFNDGVATQSVYITPEKIPQAPSPVNAACOTTLFG  
RGLGENDSPD"  
complement (6321. .7562)  
/locus\_tag="t1464"  
complement (6321. .7562)  
/locus\_tag="t1464"  
/note="corresponds to STY1517 from Accession AL513382:  
Salmonella typhi CT18"  
/codon\_start=1  
/transl\_table=11  
/product="putative multidrug efflux protein"  
/protein\_id="AAO69104.1"  
/db\_xref="GI:29137542"  
/translation="NNTNVYENTDSEITTPLNKRRILFVLLVGLVAASTAAVMSVLP

Query Match	Best Local Similarity	Score	Length
Matches 136;	Conservative 0;	Mismatches 3;	Indels 0; Gaps 0;
QY	1	GTGACTCTGGTCGACGAACCTTAATAAATGACCCCTGCTCTTTCTTCGAAAGAGG	60
DB	250187	GTGATCTGCTGCGACGAACCTTAATAAATGATGCTCACCCCTCTTTCTTCGAAAGAGG	250246
QY	61	GTGACTATTTGTCTGTTTATTAACGTGTTATCCCAAGACCAATATCAACGCTAGAC	120
DB	250247	GTGACTATTTGTCTGTTTATTAACGTGTTATCCCAAGACCAATATCAACGCTAGAC	250306

QY 121 TGTTCTATTGTTAACA 139  
 |||||  
 DB 250307 TGTTCTATTGTTAACA 250325

RESULT 6  
 STYPHOPO  
 LOCUS  
 DEFINITION S.typhimurium phoP protein and membrane protein phoQ genes,  
 complete cds.  
 STYPHOPO 2190 bp DNA linear BCT 26-APR-1993

ACCESSION M24424.1 GI:154265  
 VERSION M24424  
 KEYWORDS membrane protein; phoP gene; phoQ protein.  
 SOURCE Salmonella typhimurium  
 ORGANISM Salmonella typhimurium  
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 Enterobacteriaceae; Salmonella.

REFERENCE 1 (bases 1 to 2190)  
 Miller S.I., Kukulski A.M. and Mekalanos J.J.  
 A two-component regulatory system (phoP phoQ) controls Salmonella  
 typhimurium virulence

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 86 (13), 5054-5058 (1989)  
 MEDLINE 89296942  
 PUBMED 2544889

COMMENT Original source text: S.typhimurium (strain LT2) DNA.  
 Draft entry and printed copy of sequence for [1] kindly provided by  
 S.I. Miller, 01-MAY-1989.

FEATURES  
 source Location/Qualifiers

1..2190  
 /organism="Salmonella typhimurium"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:602"  
 141..815  
 /note="phoP protein"  
 /codon\_start=1  
 /transl\_table=1  
 /protein\_id="AAA27188.1"  
 /db\_xref="GI:154265"

CDS  
 815..2170  
 /note="membrane protein phoQ"  
 /codon\_start=1  
 /transl\_table=1  
 /protein\_id="AAA27189.1"  
 /db\_xref="GI:154267"

/translation="MNKFAHPLPLSLRVPFLATAGVYVLSIAGVAVAVGVSVF  
 DKTRFLRGESNLFTYLLAKMENNKSIVLPEPMDQRIIPWLSIQEPMKTNQPH  
 ELSTVNDATSLISEDSHQAQEKIKEYREDDDJAEWTHSAVNIYPTATAPQLTIVV  
 DTPLEKRSYVWVWFVYVLANLVLPLMTAWMSLRPIELAREVLEBHR  
 BMNPETREILTSIVNLNQLKSERERYNKRTTLTDLTHLKTPLAVLOSTLNL  
 NEMSVSKAPVWLBOISRIISOIGYVLRASMRGSGVLSREHLPVAPLNLISAL  
 NRYVQKGVNISMDSIPESIPGEQNDPVENGVNLDNAKCYLCLEPVSATQTDHL  
 HIFVEDGPGIIPHSKRSIVFDKQGRADTLRPQILASDSLLGARHEVVFGRQHPYQKE  
 E"

## ORIGIN

Query Match 92.8%; Score 129; DB 1; Length 2190;  
 Best Local Similarity 100.0%; Pred. No. 3.2e-26;  
 Matches 129; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 TCGACCACTTAATATATAGCTGCTGCCTACCTCTTTCTTCAGAAAGAGGCTGACTATTT 70  
 |||||  
 DB 1 TCGACCACTTAATATATAGCTGCTGCCTACCTCTTTCTTCAGAAAGAGGCTGACTATTT 60  
 |||||  
 QY 71 GTCTGTTTATTATACGTTTATCCCAAGACCAATATACGCTAGACTGTTCTTATT 130  
 |||||  
 DB 61 GTCTGTTTATTATACGTTTATCCCAAGACCAATATACGCTAGACTGTTCTTATT 120  
 |||||

QY 131 GTTAACACA 139  
 |||||  
 DB 121 GTTAACACA 129

RESULT 7  
 AE008754/c  
 LOCUS  
 DEFINITION Salmomella typhimurium LT2, section 58 of 220 of the complete  
 genome.  
 Salmomella typhimurium LT2, section 58 of 220 of the complete  
 genome.

ACCESSION AE008754 AE006468  
 VERSION AE008754.1 GI:16419750  
 KEYWORDS

SOURCE  
 ORGANISM Salmonella typhimurium LT2  
 Salmonella typhimurium LT2  
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 Enterobacteriaceae; Salmonella.

REFERENCE 1 (bases 1 to 24578)  
 McClelland, M., Sanderson, K.E., Spieth, J., Clifton, S.W.,  
 Laclelle, P., Courtney, L., Porwollik, S., Ali, J., Dante, M., Du, F.,  
 Hou, S., Layman, D., Leonard, S., Nguyen, C., Scott, K., Holmes, A.,  
 Grewal, N., Mulvaney, E., Ryan, R., Sun, H., Flores, L., Miller, W.,  
 Stoneking, T., Nhan, M., Waterston, R., and Wilson, R.K.  
 Complete genome sequence of Salmonella enterica serovar Typhimurium  
 LT2

JOURNAL Nature 413 (6858), 852-856 (2001)  
 MEDLINE 21534948  
 PUBMED 11677609

REFERENCE 2 (bases 1 to 24578)

AUTHORS  
 TITLE The Salmonella typhimurium Genome Sequencing Project  
 CONSRTM Direct Submission  
 JOURNAL Submitted (29-MAR-2001) Genome Sequencing Center, Department of  
 Genetics, Washington University School of Medicine, 4444 Forest  
 Park Boulevard, St. Louis, MO 63108, USA

COMMENT Supported by NIH grant 5U 01 AI43283

COMMENT  
 Coding sequences below are predicted from manually evaluated  
 computer analysis, using similarity information and the programs;  
 GLIMMER, <http://www.tigr.org/softlab/glimmer/glimmer.html> and  
 GeneMark; <http://opal.biology.gatech.edu/GeneMark/>

EC numbers were kindly provided by Junko Yabuzaki and the Kyoto  
 Encyclopedia of Genes and Genomes; <http://www.genome.ad.jp/kegg/>,  
 and Pedro Romero and Peter Karp at EcoCyc;  
<http://ecocyc.PangeaSystems.com/ecocyc/>

The analyses of ribosome binding sites and promoter binding sites  
 were kindly provided by Heladia Salgado, Julio Collado-Vides and  
 Reguonb;  
[http://kinch.cifn.unam.mx:8850/db/reguondb\\_intro.framesec](http://kinch.cifn.unam.mx:8850/db/reguondb_intro.framesec)

This sequence was finished as follows unless otherwise noted: all  
 regions were double stranded, sequenced with an alternate  
 chemistries or covered by high quality data (i.e., phred quality >=  
 30); an attempt was made to resolve all sequencing problems, such  
 as compressions and repeats: all regions were covered by sequence  
 from more than one ml3 subclone.

## FEATURES

source Location/Qualifiers

1..24578  
 /organism="Salmonella typhimurium LT2"  
 /mol\_type="genomic DNA"  
 /strain="LT2; SGGC 1412; ATCC 700720"  
 /db\_xref="ATCC:700720"  
 /db\_xref="taxon:99287"  
 /note="LT2"  
 complement(87..1472)  
 /gene="purB"  
 /note="synonym: STM1232"  
 complement(87..1457)  
 /gene="purB"  
 /EC number="4.3.2.2"  
 /note="similar to E. coli adenylsuccinate lyase

CDS

(AAC74215.1); Blastp hit to AAC74215.1 (456 aa), 94% identity in aa 1 - 456"  
/codon\_start=1  
/transl\_table=1  
/product="adenylosuccinate lyase"  
/protein\_id="AAL20161.1"  
/db\_xref="GI:16419751"  
/translation="MELSSLTAVSPVDGRYGVKVALRGIFSEYGLKIFRQVEFVRL  
OKLAHAHAKVEPAFADANGVLDITLVANFMEEDAPARKTIERPTNHDAVAFLEK  
KVALPAHDAIDSEFHFACSTEDINLISHALMTLADVEILLPYRQVIVANAQVLATQ  
VRDPLISRTHGQAPATPSTLIGKEMNAVYRREBROFRONOVEILGKINGAVGNNAH  
YAAPVDMHGSSEFVTSLSIGQNNPYTQTQIEPHYVILFPCIAKRPNTLLIDPRDWY  
GAILNHRKQKTIAGEISSTMPHKQNFIDENSEGNGLSNAYVHLANKLPYRQKQ  
RDLDTSTLYRNLGVGIGYALAYQSTLKGSKLVNRRDHLIDEDHWEVIAEPTIQTV  
MRVYIEFRYKELTKELTGKREVDAGMKQFIDSLALPEAKETRLKAMTPANTYIGRAVTL  
VDELK"  
complement(1461..2160)  
/gene="ycfc"  
/note="synonym: STM1233"  
complement(1461..2108)  
/gene="ycfc"  
/note="similar to E. coli orf, hypothetical protein  
(AAC74216.1); Blastp hit to AAC74216.1 (213 aa), 94%  
identity in aa 1 - 213"  
/codon\_start=1  
/transl\_table=1  
/product="unknown membrane associated protein"  
/protein\_id="AAL20162.1"  
/db\_xref="GI:16419752"  
/translation="MIVAKNYDITLALSGTCQSRRLVQQLAHGCHDADALHYSLNS  
VIDNMPSTLVGVGSEANLRLGLETLLGVNLASSRQGLNELTRYTSLVNLKRKLS  
SAKKGLNLTGDRINGLQRLDFLDQSTLMSNMGIVVDVISPGLRPIQVTSPPAVLV  
QSPVOAKRKASLILGPRVAVLMQVGGRLQLMFSRRRLTYKQALLAHITPEL"  
complement(1467..1472)  
/gene="ycfc"  
/note="putative RBS for purB; RegulonDB: STMS1H001488"  
complement(2112..2117)  
/gene="ycfc"  
/note="putative RBS for ycfC; RegulonDB: STMS1H001489"  
complement(2134..2142)  
/gene="ycfc"  
/note="putative -10\_signal for ycfC;  
RegulonDB: STMLTH004479"  
complement(2152..2160)  
/gene="ycfc"  
/note="putative -35\_signal for ycfC;  
RegulonDB: STMLTH004479"  
complement(2189..3349)  
/gene="trmU"  
/note="synonym: STM1234"  
complement(2189..3340)  
/gene="trmU"  
/EC\_number="2.1.1.61"  
/note="similar to E. coli orf, hypothetical protein  
(AAC74217.1); Blastp hit to AAC74217.1 (383 aa), 94%  
identity in aa 1 - 383"  
/codon\_start=1  
/transl\_table=1  
/product="trmU  
(5-methylaminomethyl-2-thiouridyate)-methyltransferase"  
/protein\_id="AAL20163.1"  
/db\_xref="GI:16419753"  
/translation="MIEYALKRNVASIPMSPPKKVIYMGSGVDSVSATLLQQQQ  
VYQYEGFPMKMBEEDGEERYCTAALADLADQACDGLKELHTVNFRAEYVNVDFLFL  
EBYEGATPMPDILCNKEIKKFAFLFEAEDGADYIATGHTVRRADNGSKRLRLG  
DGNKDDSYFLTYTLGHEIOIAQSLPEVGELEKRPVARKIAEDLVIAKKDOSTGICETG  
RKRRDITGRILPAQPGKILITVDDDEIGEHQGLMHTLGGKGLGIGTKDGEDMDPY  
VVDQVNNVLIYVAGQHEHRLMSVGLIAQQLHMVDREPTGTLRCTVTKRYQTDIPW  
TIALANDRREVI FDEPVAATPFGQAVFYSGVCLGGIIEQRLLPY"  
complement(3344..3349)  
/gene="trmU"  
/note="putative RBS for trmU; RegulonDB: STMS1H001490"  
complement(3349..3820)

```

/gene="ymfB"
/note="synonym: STM1235"
complement (3349. .3810)
/gene="ymfB"
/note="similar to E. coli putative phosphohydrolase
(AAC74218.1) ; Blastp hit to AAC74218.1 (153 aa), 84%
identity in aa 1 - 152"
/codon_start=1
/transl_table=11
/product="putative Mut-Like protein"
/protein_id="AAL20164.1"
/db_xref="GI:16419754"
/translation="MFKPHYTACIVHAEDKPLVETTINGSLMNPAGHLEADETIL
AQAARLEMBERTGITRPOQYFIRMHOMIAADKTPRLRPAITLDHICANDPHNDLID
CCWVADETLNAPNLRSPVAVSRSTQGRPLSTVSEFNPFTGKYK"
complement (3815. .3820)
/gene="ymfB"
/note="putative RBS for ymfB; RegulonDB:STMS1H001491"
complement (3822. .4159)
/gene="STM1236"
complement (3822. .4151)
/gene="STM1236"
/codon_start=1
/transl_table=11
/product="putative periplasmic protein"
/protein_id="AAL20165.1"
/db_xref="GI:16419755"
/translation="MKRLITAFALTPLSVNAASFDQQAQAADEKAI CAHLTLNDKCE
VEMHTKQFLPKGLFAMGSRGALQDAQQSWLKORQQKADVTCLTKAVNERIKQDLVIYI
DHI DDKPL"
complement (4148. .4813)
/gene="ymfC"
/note="synonym: STM1237"
complement (4148. .4813)
/gene="ymfC"
/note="similar to E. coli orf, hypothetical protein
(AAC74219.1) ; Blastp hit to AAC74219.1 (207 aa), 83%
identity in aa 1 - 205"
/codon_start=1
/transl_table=11
/product="putative ribosomal large subunit pseudouridine
synthase"
/protein_id="AAL20166.1"
/db_xref="GI:16419756"
/translation="MRQLISSEMTQKTSFRNHVKKPSSRQASKSKRENQPKRVILE
NKPEADLPQFTDEAGRRLLDFIPQGVYAAGRDRSEGLVLTNGLAQARLPQPE
KRKGIYVQVEGI PDNAALQALRTGTLNDGPTLPAGIEIVASPDMLMPPTPRIEE
KNIPSTSLKVTLYEGRRNRQVRMTAHGHPTLRILIRYSMDYTLNGDNGMRRLAQ
KDR"
complement (4154. .4159)
/gene="STM1236"
/note="putative RBS for STM1236; RegulonDB:STMS1H001492"
4740. .4754
/note="putative binding site for FruR, RegulonDB:
STMS1H00184"
/bound_moiety="FruR"
4829. .6235
/gene="icda"
/note="synonym: STM1238"
4829. 4837
/gene="icda"

gene
protein_bind
RBS
-35_signal

Query Match 81.3%; Score 113; DB 1; Length 24578;
Best Local Similarity 100.0%; Pred. No. 7.3e-22;
Matches 113; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 GTGACTGTGTCGACGAACCTTAATAATAGCCTGCTACACCTCTTTTCTTCAGAAAGG 60
113 GTGACTGTGTCGACGAACCTTAATAAGCCTGCTACACCTCTTTTCTTCAGAAAGG 54
61 GTGACTATTTGTCGGATTATTAATCGTTTATCCCAAGACCATTAATCAAC 113
53 GTGACTATTTGTCGGATTATTAATCGTTTATCCCAAGACCATTAATCAAC 1

```





Center code: MUGSC  
Web site: <http://genome.wustl.edu>  
Contact: [submissions@watson.wustl.edu](mailto:submissions@watson.wustl.edu)  
----- Summary Statistics -----  
Center project name: M\_BB0198116  
-----

## NOTICE:

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded, sequenced with an alternate  
chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by sequence  
from more than one subclone; and the assembly was confirmed by  
restriction digest.

## MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. Wes Warren,  
Department of Genetics, Washington University, St. Louis MO. For  
additional information about the map position of this sequence, see  
<http://genome.wustl.edu>

## SOURCE INFORMATION:

The RPCT-24 BAC Library has been constructed by Pieter de Jong and  
coworkers (<http://www.chori.org>) from male C57BL/6J mouse spleen  
and/or brain genomic DNA. The clone and detailed information can be  
obtained from Pieter de Jong and coworkers at <http://www.chori.org>

## NEIGHBORING SEQUENCE INFORMATION:

This sequence is the entire insert of the clone.

## Location/Qualifiers

1.166554

/organism="Mus musculus"

/mol\_type="Genomic DNA"

/db\_xref="taxon:10090"

/chromosome="10"

/map="10"

/clone="RP24-198116"

/clone\_1fb="RPCT-24"

888..940

/rpt\_family="ID"

1061..1435

/note="Unresolved simple sequence repeat."

1453..1595

/rpt\_family="Alu"

1713..1904

/rpt\_family="B2"

2261..2337

/rpt\_family="ID"

2513..2639

/rpt\_family="Alu"

3303..3393

/rpt\_family="Alu"

5002..5190

/rpt\_family="B2"

5736..5852

/rpt\_family="B2"

5853..6041

/note="Unresolved simple sequence repeat."

6043..6115

/rpt\_family="Alu"

6772..6963

/rpt\_family="B2"

6983..7246

/rpt\_family="B2"

7369..7521

/rpt\_family="B2"

8649..8816

/rpt\_family="B2"

9075..9188

/rpt\_family="Alu"

9107..9257

/rpt\_family="B2"

9283..9468

/rpt\_family="B2"

9717..9785

/rpt\_family="B2"

9848..9876

/rpt\_family="B2"

10164..10230

/rpt\_family="B2"

10513..10595

/note="Sequence derived from one plasmid subclone."

10547..10685

/rpt\_family="L2"

10691..10832

/rpt\_family="Alu"

10935..11093

/rpt\_family="L1"

11222..11456

/rpt\_family="B2"

11468..11602

/rpt\_family="B2"

11549..11813

/note="Unresolved tandem repeat."

11604..11652

/rpt\_family="B2"

11648..11710

/rpt\_family="B2"

11686..11760

/rpt\_family="B2"

11754..11810

/rpt\_family="B2"

11884..12147

/rpt\_family="L1"

12172..12378

/rpt\_family="L1"

12427..12542

/rpt\_family="Alu"

12559..12690

/rpt\_family="L1"

13153..13354

/rpt\_family="B2"

13639..13968

/rpt\_family="L1"

13969..14311

/rpt\_family="L1"

15247..15367

/rpt\_family="Alu"

16734..16796

/rpt\_family="B2"

18600..18806

/rpt\_family="B2"

19076..19282

/rpt\_family="B2"

19577..19916

/rpt\_family="B2"

21650..21932

/rpt\_family="B2"

21935..22042

/rpt\_family="Alu"

22106..22389

/rpt\_family="B2"

23099..23401

/rpt\_family="B2"

23635..23898

/rpt\_family="B2"

24718..25099

/rpt\_family="B2"

25134..25197

/rpt\_family="ID"

25386..25462

/rpt\_family="ID"

25565..25678

/rpt\_family="B2"

FEATURES	source
JOURNAL	21929760
MEDLINE	11932238
PUBMED	2 (bases 1 to 11716)
REFERENCE	Birken, B.
AUTHORS	Direct Submission
TITLE	Submitted (20-MAR-2002) Center for Genome Research, Whitehead Institute, Nine Cambridge Center, Cambridge, MA 02141, USA
JOURNAL	location/Qualifiers
FEATURES	1. 11716
source	/organism="Methanosarcina acetivorans C2A"
gene	/mol_type="genomic DNA"
CDS	/strain="C2A"
	/db_xref="taxon:188937"
	complement(648..1085)
	/gene="MA1763"
	complement(648..1085)
	/gene="MA1763"

/codon\_start=1  
 /transl\_table=1  
 /product="transcriptional regulator, Cpgc family"  
 /protein\_id="AA05169.1"  
 /db\_xref="GI:19915654"  
 /translation="MSGDTMEAGLMRIGVSIIPDALLTQFDEIIKKRDSRSEAIRDA  
 LISITYEWMEDIKGRVGTIAVIYDHKSGLSNLSINQHHSHLIKYSVMHYLDT  
 DDCBELIVLQNGOEITELAGSIIALKGVFSKLTITYDPKKI"  
 complement(12087. .4807)  
 /gene="MA1764"  
 /complement(2087. .4807)  
 /gene="MA1764"  
 /codon\_start=1  
 /transl\_table=1  
 /product="cell surface protein"  
 /protein\_id="AA05170.1"  
 /db\_xref="GI:19915655"  
 /translation="MEFRFMKLTRKFFGNLRLSLAVLLFSLVCTAGEBLYFEBO  
 EAVSDGSVASLILYDQAPSLGAKNLSTISDPVAITGVFEFPMAGINNSSL  
 PADLVQKADILREKINAGLDVPLVTSGLRAGLRLVTVDRDLSNRTISYV  
 VREALTSGTTSSENPVSIIPABSTLSSGASQYRELRADFTGIGISYDPLLEMA  
 VGEKTVSEPSWALSENSSNPNTSIKVKAVDNDVAKADENVLAVTTFAETPGE  
 SEISITFRLRDDSGSNINAVAEASVSQVSGSTLPAAKTAFTPGDSSTLVQFDE  
 STGEPTATWPFQDGAATEQSPHNTYSGNTVNLVTNDDGSGSEVKTDITVSE  
 STPEEPVPAFTADYNTGVPLTVNFTQOSTEAPFSMAEDPNDGTVDSTQONPSY  
 TYTAGKTVNLVTANAEGSDSEVKIDITYVSESSPTPEEPVPAFTADYNTGVPLT  
 VNFQDTSGSTSLMDPNDGTSAETQONPHNTNAGTYVNLTVISEGNSRYVAD  
 YITVSESTPEEPVPAFTADYNTGVPLTVNFTDOSTMPTSMAMPDNDGMST  
 EQNSYETTAGENTVNLTVSEVSDGSEVKEVETVYTDSTTTPEARPDILVSIILP  
 SSIATNTECNTGINTSGESDAGTPAALTSVNAVVDTHAVSGIVSGSATAVNFVK  
 PVAGVSVKVSADAGNVIASEDTNNELDLCIVBEASGSGOSONGSSSVSLNVTIV  
 PVCEVSPSALKFGEISAGOTSEPOLTLTKNCVSGIKYTAESDSSAADNIFGLGL  
 LLDQIMSDYCNVIGNSQENATLALHVPDSYDTGNKKGITTPMAEVA"  
 complement(4828. .5748)  
 /gene="MA1765"  
 /complement(4828. .5748)  
 /gene="MA1765"  
 /codon\_start=1  
 /transl\_table=1  
 /product="conserved hypothetical protein"  
 /protein\_id="AA05171.1"  
 /db\_xref="GI:19915656"  
 /translation="MOKRSGVILSILAPATLSALICTGLAEASPILLIIPSEISL  
 GTGEICELSLSDSLPBGLSYKTVLELTPDVALEIQAANPEKASLTVDSELPASSV  
 KLRADVLDEALDGEAEDELVALTLTKLBGGSTGITYVTSMPDSSERTYILBEAEP  
 KEDGSDSDGDESDGDEBDEBENTVEDNTEVTSKTSQPSQSSNSVSIQDPP  
 NETAEKMSQADENSLSLKEVEIISSEPSKQDPLLHSDEADESKNDTGNTGTGNS  
 TGNSTLPGRGFYSIVFVFVYISRSKK"  
 complement(5861. .11104)  
 /gene="MA1766"  
 /complement(5861. .11104)  
 /gene="MA1766"  
 /codon\_start=1  
 /transl\_table=1  
 /product="cell surface protein"  
 /protein\_id="AA05172.1"  
 /db\_xref="GI:19915657"  
 /translation="MDPCYLLMWELVHIALGFELYRGTGKVTDRKLIIVLLTLILIC  
 SSVGLTLFDDAAADENVGGLPLTVQETIEYTGOLMPEATPAPMKWGCVTKTLPBE  
 VAAVRGPIAAVRALISAYCGHOSDVAFTLSMDGNDPQAVQWENPCHGAFQYIL  
 IDPNTWETLQDNSEFQDGHGTQOEPIYMLNHNENVTQSPFMWYKOLIGQITVNNQ  
 DTTGVSIDGKIKITLVANINDSSTYOTTYYNREHGYCSIFEDNGEAAVNGTSGFG  
 TTGLSEVSALTVDVYANNGCGEFTPADNNEFYTGTSBEGTFVNLQDLDRDPDTQ  
 GAYSGIDSNNVSTVSGSDVTLGYARFGLDEGTAAPFKLPLAFLVKSPIETSSQPP  
 VADFTADYVSGDAPLEVQFTDASTGTVSSYAMDPDNDGTVDSTOONQVYISTSGY  
 VNLVTNNEGSGSELKTDITYTVOAGVATNTLISGIVNTVPASAVFAETNFKYLV  
 NYQNTGATLNTNLSIAVASYDSGTVVNTDLSLAGLAKTIVTLIDPIRLDEGGGIL  
 TYTYRAVDPNULIETDEPTNNKSSAKPLRNYGCKKGIYMEGNSITMHPEDLIC  
 GNULYSYQDSYQAQVNGESRTETWASDLVPPDSTIEKAPLYVATNMQDTPGCYPR  
 LAINRNGTLDNGNISTGNGTLWRDMSVAGVANYEGLCVYDVTDFRSSGNSLWMT  
 PIGENICMLYPTLIVVYGNDETETKQIFINECDLGLSASVGTTPBEATTAAPFV  
 GSIDEVKTNMLYSFAGSAGDEGNLTLPINENIYATANAGSSNGSPLEAFADATNYI  
 NATNGEAGVQSTTSGMALQOILVIEYESAPSPADVDTATPTSGDAFLAVNFTDPI



```

repeat_region      5466..5853
                    /rpt_family="L2"
repeat_region      5884..6983
                    /rpt_family="L1"
repeat_region      7909..7936
                    /rpt_family="(TTTA)n"
repeat_region      7939..8944
                    /rpt_family="L1"
repeat_region      9538..9572
                    /rpt_family="AT_rich"
repeat_region      11515..11638
                    /rpt_family="MIR"
repeat_region      11820..12393
                    /rpt_family="L1"
repeat_region      12686..12756
                    /rpt_family="Alu"
repeat_region      12758..12888
                    /rpt_family="L1"
repeat_region      13182..13413
                    /rpt_family="L1"
repeat_region      18096..18137
                    /rpt_family="AT_rich"
repeat_region      18540..18883
                    /rpt_family="ERV1"
repeat_region      18954..19004
                    /rpt_family="(CATATA)n"
repeat_region      19083..19203
                    /rpt_family="MER1_type"
repeat_region      19247..19288
                    /rpt_family="(CA)n"
repeat_region      19455..19488
                    /rpt_family="AT_rich"
repeat_region      19709..19755
                    /rpt_family="AT_rich"
repeat_region      20114..20167
                    /rpt_family="L2"
repeat_region      20467..20833
                    /rpt_family="MalR"
repeat_region      20876..21032
                    /rpt_family="L2"
repeat_region      21612..21887
                    /rpt_family="Alu"
repeat_region      21893..22075
                    /rpt_family="L1"
repeat_region      22438..23791
                    /rpt_family="L1"
repeat_region      23792..24476
                    /rpt_family="L1"
misc_feature       24594..24960
                    /note="match to EST B1458889 (NID:g15249545)"
repeat_region      24660..24772
                    /rpt_family="MIR"
repeat_region      25939..26066
                    /rpt_family="MIR"
repeat_region      26474..26975
                    /rpt_family="MalR"
misc_feature       26615..26806
                    /note="match to EST B1458889 (NID:g15249545)"
repeat_region      27142..27187
                    /rpt_family="AT_rich"
repeat_region      27569..27894
                    /rpt_family="Alu"
repeat_region      29779..29801
                    /rpt_family="(TA)n"
repeat_region      29802..30127
                    /rpt_family="MER2_type"
repeat_region      30128..30184
                    /rpt_family="(TA)n"
repeat_region      30716..31118
                    /rpt_family="ERV1"
repeat_region      31162..31216
                    /rpt_family="ERV1"
repeat_region      31353..31388

```

```

repeat_region      31513..31702
                    /rpt_family="MER1_type"
repeat_region      31735..31758
                    /rpt_family="AT_rich"
repeat_region      31768..32055
                    /rpt_family="Alu"
repeat_region      32493..32513
                    /rpt_family="(TA)n"
repeat_region      33000..33140
                    /rpt_family="MIR"
repeat_region      33201..33512
                    /rpt_family="Alu"

Query Match      25.9%; Score 36; DB 9; Length 44532;
Best Local Similarity 70.6%; Pred. No. 4.8;
Matches 48; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY      39  CCCTCTTTCTCAGAAAGGGTGACTATTTGCTGCTTTATTAAGTATTATCCCA 98
          |||||
Db       13898  CTCTCTTTCTCTTATTAGTTAGTACGTAAGTGTCTGTTTATTTTTCCTAA 13839
          |||||
QY      99  AGCACCAT 106
          |||||
Db       13838  ACCACCTT 13831

RESULT 13
AC103069/c
LOCUS
DEFINITION
Rattus norvegicus clone CH230-171118. *** SEQUENCING IN PROGRESS
*** 7 unordered pieces.
AC103069
AC103069.5 GI:30579796
HTG; HTGS PHASE1; HTGS DRAFT; HTGS ENRICHED.
KEYWORDS
SOURCE
Rattus norvegicus
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 245108)
Muzny,D,Marie., Metzker,M, Lee., Abramson,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Albrooks,S., Amla,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benhmed,F.,
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Butrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Caesar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Crete,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denison,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Davies,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,T., Foster,P.,
Frazer,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,M., Guevara,W.,
Gunaratne,P., Haaland,W., Hamill,C., Hamilton,N., Hernandez,J.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Huijck,S., Hume,J., Idlebird,D., Hogue,M.,
Hollins,B., Howells,S., Huijck,S., Hume,J., Johnson,B., Johnson,R., Jolivet,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kover,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorenzow,L., Louised,H., Lozada,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindaratne,M., Mahoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhinney,S., McLeod,M.P., McNeill,T.Z., Neenen,E.,
Milosavljevic,A., Miner,G., Minja,B., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munillas,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwaokemele,O., Okwuonu,G., Olarnpungoon,A., Pal,S., Parks,K.,

```

	REFERENCE	Pasternak,S., Paul,H., Perez,L., Perez,L., Pfannkuch,C., Piopier,F., Polidexter,A., Popovic,D., Primus,E., Pu,L.-L., Piazzi,M., Quirroz,J., Rachlin,B., Reeves,K., Regier,M.A., Reigh,R., Rellly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rivers,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajd,D., Speed,A., Sodergren,E., Song,X.-Z., Sorrell,R., Soza,J., Steinle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Ueman,K., Valas,R., Varga,V., Villaseana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,R., Warren,R., Wei,X., White,F., Williams,D., Willison,R., Wleczyk,R., Wooden,H., Morley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Weidenstock,G., A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinerstock,G., A., Gibbs,R.A.
TITLE	JOURNAL	Unpublished
REFERENCE	AUTHORS	Morley,K.C.
TITLE	JOURNAL	Submitted (24-NOV-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE	AUTHORS	Rat Genome Sequencing Consortium.
TITLE	JOURNAL	Submitted (13-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT		On May 13, 2003 this sequence version replaced gi:3326957. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas ( <a href="http://www.hgsc.bcm.tmc.edu/projects/rat/">http://www.hgsc.bcm.tmc.edu/projects/rat/</a> ). Each contig described in the feature table below represents a scaffold in the Atlas assembly ('a' contig-scaffold). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.
		----- Genome Center Center: Baylor College of Medicine Center code: BCM Web site: <a href="http://www.hgsc.bcm.tmc.edu/">http://www.hgsc.bcm.tmc.edu/</a> Contact: hgsc-help@bcm.tmc.edu ----- Project Information Center project name: GJDK Center Clone name: CH230-171118 ----- Summary Statistics Assembly program: Atlas 3.0; Consensus quality: 210210 bases at least Q40 Consensus quality: 214318 bases at least Q30 Consensus quality: 217468 bases at least Q20 Estimated insert size: 221660; sum-of-contigs estimation Quality coverage: 5x in Q20 bases; sum-of-contigs estimation  NOTE: Estimated insert size may differ from sequence length (see <a href="http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html">http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html</a> ). NOTE: This is a 'working draft' sequence. It currently consists of 7 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.  1      4120: contig of 4120 bp in length *      4121    gap of unknown length *      230259: contig of 226039 bp in length

Menghinley, B., Mapus, P., Martin, K., Martin, R., Martinez, E.,  
Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,  
Milosavljevic, A., Miner, G., Minia, E., Montemayor, J., Moore, S.,  
Moragan, M., Morris, K., Morris, S., Naidasa, M., Murphy, M., Neir, L.,  
Nervetti, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,  
Nowoklehen, O., Okunolu, G., Olarnpusagoon, A., Pal, S., Parks, K.,  
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,  
Plummer, F., Poldrester, A., Popovic, D., Primus, E., Pu, L.-L.,  
Puzzo, M., Quiroz, J., Rachlin, B., Reeves, K., Regier, M.A., Reigh, R.,  
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,  
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.,  
Sanders, W., Savary, G., Scherer, S., Scott, G., Shamsan, S., Shen, H.,  
Shetty, J., Shvartsbeyn, A., Slason, I., Sitter, C.D., Smajd, D.,  
Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,  
Steinle, M., Strong, R., Sutton, A., Svalke, A., Taber, P., Taylor, C.,  
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmant, K.,  
Valas, R., Vera, V., Villaseca, D., Waldron, L., Walker, B., Wang, J.,  
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,  
Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K.,  
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,  
Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von  
Nedehausen, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,  
Weinstock, G. and Gibbs, R.A.

Direct Submission

Unpublished

2 (bases 1 to 162561)

Worley, K.C.

Direct Submission

Submitted (05-MAR-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 162561)

Rat Genome Sequencing Consortium.

Direct Submission

Submitted (20-NOV-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

On Nov 20, 2002 this sequence version replaced gi:22856569.  
The sequence in this assembly is a combination of BAC based reads  
and whole genome shotgun sequencing reads assembled using Atlas  
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described  
in the feature table below represents a scaffold in the Atlas  
assembly (a 'contig-scaffold'). Within each contig-scaffold,  
individual sequence contigs are ordered and oriented, and separated  
by sized gaps filled with Ns to the estimated size. The sequence  
may extend beyond the ends of the clone and there may be sequence  
contigs within a contig-scaffold that consist entirely of whole  
genome shotgun sequence reads. Both end sequences and whole genome  
shotgun sequence only contigs will be indicated in the feature  
table.

----- Genome Center -----

Center: Baylor College of Medicine

Center code: BCM

Web site: http://www.hgsc.bcm.tmc.edu/  
Contact: hgsc-help@bcm.tmc.edu

----- Project Information -----

Project name: GPAA

Center project name: GPAA

Center clone name: CH230-330F6

----- Summary Statistics -----

Assembly program: Phrap; version 0.990329

Consensus quality: 146385 bases at least Q40

Consensus quality: 147074 bases at least Q30

Consensus quality: 147633 bases at least Q20

Estimated insert size: 150941; sum-of-contigs estimation

Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

-----

\* NOTE: Estimated insert size may differ from sequence length  
\* (see http://www.hgsc.bcm.tmc.edu/docs/genbank\_draft\_data.html).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 3 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.

FEATURES  
source location/Qualifiers  
1. 126477  
/organism="Medicago truncatula"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:3880"  
/clone="mch2-17n5"  
/clone\_lib="Medicago truncatula BAC library H2"

ORIGIN

Query Match 25.2%; Score 35; DB 8; Length 126477;  
Best Local Similarity 63.9%; Pred. No. 7.8;  
Matches 53; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

QY 55 AAGAGGGTACTATTGCTGTTTATTACTGTTATCCCAAGCACCATTATCAACG 114  
DB 118152 AAGAGGTGATTATTAGCCAATTACACACATATATCCCAAAATGACATCATCATG 118093

QY 115 CTGACTGTTCTTATTGTTAACA 137  
DB 118092 TTATCTATATTCAAAAGACAA 118070

Search completed: August 3, 2005, 23:49:12  
Job time : 6699 secs